
Université Pierre et Marie Curie
Master de Mathématique
2005-2006

Probabilités Approfondies

Polycopié: J. Lacroix & P. Priouret, Cours: J. Lacroix

Université Pierre et Marie Curie
Master de Mathématiques et Applications
Année 2005/2006

Probabilités Approfondies

Jean Lacroix & Pierre Priouret

Mode d'emploi

Ce polycopié est destiné aux étudiants de l'U.E. "Probabilités Approfondies" du Master de Mathématiques de l'Université Pierre et Marie Curie. En principe il s'adresse donc à des étudiants ayant suivi un cours d'intégration et un premier cours de probabilités. Cependant le chapitre 1 contient un rappel de tous les résultats d'intégration utilisés par la suite. Quant au chapitre 2 qui introduit les principales notions de probabilités, il est relativement autonome et peut éventuellement être abordé par un étudiant n'ayant jamais suivi de cours de probabilités. Le chapitre 3 présente les espérances conditionnelles et le calcul des lois conditionnelles. Les chapitres 4 et 5 sont consacrés aux deux sujets essentiels de ce module, d'une part l'étude des chaînes de Markov à temps discret et à valeurs dénombrables et d'autre part, l'étude des martingales à temps discret.
Un certain nombre de résultats, figurant classiquement dans un cours de probabilités au niveau maîtrise, ont étés rejetés en annexe, car ils ne sont pas vraiment nécessaires pour la compréhension des deux chapitres principaux, à savoir les chapitres 4 et 5. Il est néanmoins vivement recommandé au lecteur d'en prendre connaissance.
Ce polycopié est divisé en chapitres, sections et sous-sections. Ainsi 3.2.4 renvoie au chapitre 3, section 2, sous-section 4 et 5.4 renvoie chapitre 5, section 4. A l'intérieur d'une même section, les énoncés sont numérotés en continu. Ainsi "d'après le th. 5.4.6" renvoie au chapitre 5, section 4, énoncé 6. Quant aux égalités, elles sont numérotées entre parenthèses et en continu au sein d'un même chapitre. Ainsi "vu (3.5)" réfère à la cinquième égalité numérotée du chapitre 3. Le signe " $\blacksquare$ " indique la fin d'une preuve.
Enfin une courte bibliographie présente quelques ouvrages de base sur le sujet ainsi que quelques textes d'exercices corrigés.

Table des matières

Chapitre 1

Rappels d'intégration

Dans ce premier chapitre, on rappelle les principaux résultats de la théorie de la mesure et de l'intégration puis on étudie de façon plus détaillée les mesures bornées sur $\mathbb{R}^d$ (convolution, transformation de Fourier, convergences).

1.1. Tribus

Soient E un ensemble et $\mathcal{B} \subset \mathcal{P}(E)$. On dit que $\mathcal{B}$ est une algèbre (resp. une tribu) si $E \in \mathcal{B}$, si $\mathcal{B}$ est stable par passage au complémentaire et par réunion et intersection finies (resp. dénombrables). Un couple $(E, \mathcal{B})$, $\mathcal{B}$ tribu sur E, s'appelle un espace mesurable. S'il est souvent possible de décrire les éléments d'une algèbre, il n'en est pas de même pour ceux d'une tribu. On remarque que $\mathcal{P}(E)$ est une tribu et que l'intersection d'une famille quelconque de tribus est une tribu. Donc, étant donné $\mathcal{C} \subset \mathcal{P}(E)$, on peut considérer la plus petite tribu contenant $\mathcal{C}$, c'est l'intersection de toutes les tribus contenant $\mathcal{C}$. Cette tribu se note $\sigma(\mathcal{C})$ et s'appelle la tribu engendrée par $\mathcal{C}$. Le résultat suivant, appelé théorème de classe monotone, sera d'un usage constant dans la suite.

Proposition 1.1.1. *Soient $\mathcal{C} \subset \mathcal{M} \subset \mathcal{P}(E)$. On suppose que $\mathcal{C}$ est stable par intersection finie, que $E \in \mathcal{M}$, que $A, B \in \mathcal{M}$ et $A \subset B$ impliquent $B \setminus A \in \mathcal{M}$ et que $\mathcal{M}$ est stable par limite croissante. Alors $\sigma(\mathcal{C}) \subset \mathcal{M}$.*

Supposons $E = \mathbb{R}^d$ et soit $\mathcal{O}$ la classe des ouverts de E. La tribu $\sigma(\mathcal{O})$ s'appelle la tribu borélienne de $\mathbb{R}^d$ et se note $\mathcal{B}(\mathbb{R}^d)$. Il est facile de voir qu'elle est aussi engendrée par les fermés, par les boules, par les pavés et même par les pavés à coordonnées rationnelles (cette dernière famille ayant l'avantage d'être dénombrable). Si $d = 1$, on considérera, outre $\mathcal{B}(\mathbb{R})$, $\mathcal{B}(\mathbb{R}^+) = \{A \in \mathcal{B}(\mathbb{R}), A \subset \mathbb{R}^+\}$, $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\mathcal{B}(\mathbb{R}), \{+\infty\}, \{-\infty\})$ et $\mathcal{B}(\overline{\mathbb{R}}^+) = \sigma(\mathcal{B}(\mathbb{R}^+), \{+\infty\})$. On étend les opérations usuelles à $\overline{\mathbb{R}}^+$ en posant $(+\infty) \times 0 = 0 \times (+\infty) = 0$.

Soient $(E_1, \mathcal{B}_1)$ et $(E_2, \mathcal{B}_2)$ deux espaces mesurables. Une application f de E_1 dans E_2 est dite mesurable si, pour tout $A \in \mathcal{B}_2$, $f^{-1}(A) \in \mathcal{B}_1$. Il est facile de voir que pour cela, il suffit que $f^{-1}(A) \in \mathcal{B}_1$ pour tout $A \in \mathcal{C}$ avec $\sigma(\mathcal{C}) = \mathcal{B}_2$. Ceci implique que,

si f est continue de $\mathbb{R}^d$ dans $\mathbb{R}^m$, f est mesurable pour les tribus boréliennes (on dit alors que f est borélienne). De plus, cette notion est transitive i.e. la composée de deux applications mesurables est mesurable. Quand l'espace d'arrivée est $\mathbb{R}$, $\overline{\mathbb{R}}$, $\overline{\mathbb{R}}^+$, $\mathbb{R}^d$, $\mathbb{C}$, il est toujours supposé muni de sa tribu borélienne.

Soit $(E, \mathcal{B})$ un espace mesurable. Pour qu'une application numérique soit mesurable, il suffit que, pour tout $a \in \mathbb{R}$, $\{f > a\} := \{x,\ f(x) > a\} \in \mathcal{B}$. On peut aussi considérer $\{f < a\}$, $\{f \leq a\}$, $\{f \geq a\}$. Ceci implique que, si f, g, f_n sont des fonctions numériques mesurables, il en est de même de $-f$, $\sup(f, g)$, $\inf(f, g)$, $f^+ = \sup(f, 0)$, $f^- = \sup(-f, 0)$, $\sup f_n$, $\inf f_n$, $\overline{\lim}\, f_n$, $\underline{\lim}\, f_n$, $\lim f_n$ si elle existe.
Rappelons que, notant $f_n \uparrow f$ (resp.$f_n \downarrow f$) si, pour tout $x \in E$, $f_n(x)$ croît (resp. décroît) vers $f(x)$,

$$\overline{\lim}\, f_n(x) = \lim_n \downarrow \sup_{k\geq n} f_k(x), \quad \underline{\lim}\, f_n(x) = \lim_n \uparrow \inf_{k\geq n} f_k(x), \tag{1.1}$$

ces quantités étant à valeurs $\overline{\mathbb{R}}$ et que $f = \lim f_n$ ssi $\overline{\lim}\, f_n = \underline{\lim}\, f_n = f$.
Soient f, g des fonctions numériques mesurables. Alors $\phi : x \mapsto (f(x), g(x))$ est mesurable de $(E, \mathcal{B})$ dans $\mathbb{R}^2$ puisque $\phi^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B)$. Ceci implique que, si H est une application borélienne de $\mathbb{R}^2$ dans $\mathbb{R}$, $H(f, g)$ est mesurable. On en déduit que $f + g$, fg, $\frac{f}{g}$, si elle existe, sont mesurables.

Pour $A \subset B$, on appelle fonction indicatrice de A et on note $\mathbb{1}_{\{A\}}$ la fonction valant 1 sur A et 0 sur A^c (on note A^c le complémentaire de A). On a

$$\mathbb{1}_{\{A^c\}} = 1 - \mathbb{1}_{\{A\}}, \quad \mathbb{1}_{\{\cap A_n\}} = \prod_n \mathbb{1}_{\{A_n\}} = \inf \mathbb{1}_{\{A_n\}}, \quad \mathbb{1}_{\{\cup A_n\}} = \sup \mathbb{1}_{\{A_n\}}.$$

Une application f de E muni de la tribu $\mathcal{B}$ dans $\mathbb{R}$ est dite étagée si elle s'écrit $f = \sum_{k=1}^n a_k \mathbb{1}_{\{A_k\}}$, $A_k \in \mathcal{B}$. On notera:

- $\mathcal{B}_r$ l'ensemble des fonctions réelles $\mathcal{B}$-mesurables,
- $\mathcal{B}_b$ l'ensemble des fonctions réelles $\mathcal{B}$-mesurables bornées,
- $\mathcal{B}^+$ l'ensemble des fonctions $\mathcal{B}$-mesurables à valeurs $\overline{\mathbb{R}}^+$,
- $\mathcal{B}_e^+$ l'ensemble des fonctions étagées positives.

Le résultat suivant est à la base de la construction de l'intégrale

Proposition 1.1.2. *Toute $f \in \mathcal{B}^+$ est limite d'une suite croissante de fonctions de $\mathcal{B}_e^+$.*

`Preuve:`
Il suffit de considérer

$$f_n(x) = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{1}_{\{\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\}} + n \mathbb{1}_{\{f(x)\geq n\}} \blacksquare \tag{1.2}$$

Soit f une application de E dans un espace mesurable $(A, \mathcal{A})$. On note $\sigma(f)$ et on appelle tribu engendrée par f la plus petite tribu sur E rendant f mesurable. On a donc $\sigma(f) = \{f^{-1}(A),\ A \in \mathcal{A}\}$. Plus généralement si $(f_i,\, i \in I)$ est une famille d'applications de E dans des espaces mesurables $(F_i, \mathcal{F}_i)$, on note $\sigma(f_i,\, i \in I)$ et on appelle tribu engendrée par les f_i la plus petite tribu sur E rendant toutes les f_i mesurables. On a donc

$$\sigma(f_i,\, i \in I) = \sigma(f_i^{-1}(A_i),\, A_i \in \mathcal{F}_i,\, i \in I).$$

On peut aussi donner une version fonctionnelle du théorème des classes monotones (prop.1.1.1):

Théorème 1.1.3. *Soient $\mathcal{H}$ un espace vectoriel de fonctions réelles bornées définies sur Ω et $\mathcal{C}$ un ensemble de parties de Ω stable par intersection finie. On suppose que,*

- $1 \in \mathcal{H}$,
- *si $f_n \in \mathcal{H}$ et si $0 \leq f_n \uparrow f$ bornée, $f \in \mathcal{H}$.*
- *pour tout $A \in \mathcal{C}$, $\mathbb{1}_{\{A\}} \in \mathcal{H}$.*

Alors $\mathcal{H}$ contient toutes les fonctions $\sigma(\mathcal{C})$-mesurables bornées.

`Preuve:`
Soit $\mathcal{M} = \{A,\ \mathbb{1}_{\{A\}} \in \mathcal{H}\}$. On a $\mathcal{C} \subset \mathcal{M}$ et, vu les hypothèses sur $\mathcal{H}$, on peut appliquer la prop.1.1.1. Donc $\sigma(\mathcal{C}) \subset \mathcal{M}$. Ceci implique que, si f est étagée sur $(\Omega, \sigma(\mathcal{C}))$, $f \in \mathcal{H}$. C'est encore vraie (prop.1.1.2) par passage à la limite croissante si f est positive bornée $\sigma(\mathcal{C})$-mesurable puis, par différence, pour toute f bornée $\sigma(\mathcal{C})$-mesurable ∎

Proposition 1.1.4. *Soit f, une application de E dans un espace mesurable $(F, \mathcal{F})$ et $h : E \to \mathbb{R}$ (resp. $E \to \overline{\mathbb{R}}^+$). Alors h est $\sigma(f)$-mesurable ssi il existe $g \in \mathcal{F}_r$ (resp. $g \in \mathcal{F}^+$) tel que $h = g \circ f$.*

`Preuve:`
Evidemment si $h = g \circ f$ alors h est $\sigma(f)$-mesurable (composition des applications mesurables). Réciproquement on pose

$$\mathcal{H} = \{\varphi\ :\ E \to \mathbb{R},\ \varphi = \psi \circ f\ ,\ \psi \in \mathcal{F}_b\}$$

On vérifie facilement que $\mathcal{H}$ est un espace vectoriel de fonctions bornées sur E vérifiant les conditions du théorème 1.1.3 avec $\mathcal{C} = \sigma(f)$ et par conséquent $\mathcal{H}$ contient toutes les fonctions $\sigma(f)$ mesurables bornées. On conclut en considérant d'abord des fonctions h bornées. ∎

1.2. Mesures

Soient I un ensemble dénombrable et $(a_i,\, i \in I)$ une famille d'éléments de $\overline{\mathbb{R}}^+$. On veut définir $\sum_{i\in I} a_i$. Soit ϕ une énumération de I i.e. une bijection de $\mathbb{N}$ sur I. On pose $S_n^\phi = \sum_{k=0}^n a_{\phi(k)}$. Evidemment S_n^ϕ croît avec n et $S^\phi = \lim \uparrow S_n^\phi$ existe dans $\overline{\mathbb{R}}^+$. Si ψ est une autre énumération de I, on a , pour n fixé et m assez grand, $\{a_{\phi(0)}, \ldots, a_{\phi(n)}\} \subset \{a_{\psi(0)}, \ldots, a_{\psi(m)}\}$, d'où $S_n^\phi \le S_m^\psi$ et $S^\phi \le S^\psi$. Permutant ϕ et ψ, on a $S^\psi \le S^\phi$ et $S^\phi = S^\psi$. On pose donc $\sum_{i\in I} a_i := \lim \uparrow S_n^\phi$, quantité qui ne dépend pas de l'énumération ϕ. Evidemment si, pour tout $i \in I$, $0 \le a_i \le b_i$, $\sum_{i\in I} a_i \le \sum_{i\in I} b_i$. On a aussi (sommation par paquets):

Théorème 1.2.1. *Soient $(a_i,\, i \in I)$ une famille d'éléments de $\overline{\mathbb{R}}^+$ et $(A_j,\, j \in J)$ une partition de I. Alors*

$$\sum_{i\in I} a_i = \sum_{j\in J} (\sum_{i\in I_j} a_i).$$

Considérons maintenant une famille $(a_i,\, i \in I)$ d'éléments de $\mathbb{C}$. On dit que cette famille est absolument sommable si $\sum_{i\in I} |a_i| < +\infty$. Dans ce cas, en décomposant la partie réelle et la partie imaginaire de a_i en leur parties positives et négatives, on voit facilement que $\sum_{i\in I} a_i := \lim S_n^\phi$ existe et est indépendante de ϕ et que le th.1.2.1 reste valable.

Définition 1.2.2. *Soit $(E, \mathcal{B})$ un espace mesurable. On appelle mesure sur $(E, \mathcal{B})$ toute application μ de $\mathcal{B}$ dans $\overline{\mathbb{R}}^+$ telle que*

- *(i) $\mu(\emptyset) = 0$,*
- *(ii) pour tous $A_n \in \mathcal{B}$ deux à deux disjoints, $\mu(\cup_n A_n) = \sum_n \mu(A_n)$.*

Le triplet $(E, \mathcal{B}, \mu)$ s'appelle un espace mesuré.

Propriétés:

- (i) si $A, B \in \mathcal{B}$ et $A \subset B$, $\mu(A) \le \mu(B)$,
- (ii) si $A_n \in \mathcal{B}$, $\mu(\cup_n A_n) \le \sum_n \mu(A_n)$,
- (iii) si $A_n \in \mathcal{B}$ et si $A_n \uparrow A$ (i.e. $\mathbb{1}_{\{A_n\}} \uparrow \mathbb{1}_{\{A\}}$), $\mu(A_n) \uparrow \mu(A)$,
- (iv) si $A_n \in \mathcal{B}$, si $A_n \downarrow A$ (i.e. $\mathbb{1}_{\{A_n\}} \downarrow \mathbb{1}_{\{A\}}$) et si, pour un n_0, $\mu(A_{n_0}) < +\infty$, $\mu(A_n) \downarrow \mu(A)$.

Si $E = \cup_n E_n$ avec $E_n \in \mathcal{B}$ et $\mu(E_n) < +\infty$, la mesure μ est dite σ-finie. Si $\mu(E) < +\infty$, la mesure μ est dite bornée. Si $\mu(E) = 1$, la mesure μ est appelée une probabilité.

Remarque. La propriété (ii) de la def.1.2.2 s'appelle σ-additivité. Si dans la def.1.2.2, on suppose que $\mathcal{B}$ est seulement une algèbre, la définition a encore un sens en rajoutant dans (ii) la condition $\cup_n A_n \in \mathcal{B}$. On a ainsi la notion de mesure sur une algèbre.

Proposition 1.2.3. *Soient μ et ν deux mesures sur $(E,\mathcal{B})$ et $\mathcal{C} \subset \mathcal{B}$ une classe d'ensembles stable par intersection finie. On suppose que, pour tout $A \in \mathcal{C}$, $\mu(A) = \nu(A) < +\infty$ et que $E = \lim \uparrow E_n$ avec $E_n \in \mathcal{C}$. Alors $\mu(A) = \nu(A)$ pour tout $A \in \sigma(\mathcal{C})$.*

`Preuve:`
Supposons d'abord $\mu(E) = \nu(E) < +\infty$. Soit $\mathcal{M} = \{A \in \mathcal{B},\ \mu(A) = \nu(A)\}$. On vérifie immédiatement que les hypothèses de la prop.1.1.2 sont vérifiées. On a donc $\sigma(\mathcal{C}) \subset \mathcal{M}$. Le cas général se traite en appliquant ce résultat aux mesures $\mu_n(A) = \mu(A \cap E_n)$ et $\nu_n(A) = \nu(A \cap E_n)$ ∎

Soit $(E,\mathcal{B},\mu)$ un espace mesuré. Un sous-ensemble A de E est dit négligeable (ou μ-négligeable s'il y a ambiguïté) si $A \subset B$ avec $B \in \mathcal{B}$ et $\mu(B) = 0$. Une propriété est vraie presque partout (en abrégé p.p.) si elle est vraie en dehors d'un ensemble négligeable. Par exemple $f = g$ p.p. signifie que $\{x \in E,\ f(x) \neq g(x)\}$ est négligeable. Si μ est une probabilité, on dit presque sûrement (en abrégé p.s.) pour presque partout. On note $\mathcal{N}$ la classe des ensembles négligeables. Il faut noter que si $A_n \in \mathcal{N}$, on a $\cup_n A_n \in \mathcal{N}$. Si $\mathcal{N} \subset \mathcal{B}$, l'espace mesuré $(E,\mathcal{B},\mu)$ est dit complet. Si ce n'est pas le cas, on peut le "compléter" de la façon suivante. On définit $\overline{\mathcal{B}} = \sigma(\mathcal{B},\mathcal{N})$. Alors $A \in \overline{\mathcal{B}}$ ssi $A = B \cup N$ avec $B \in \mathcal{B}$ et $N \in \mathcal{N}$. On peut prolonger μ à $\overline{\mathcal{B}}$ en posant $\mu(A) = \mu(B)$ (il est facile de voir que ceci ne dépend pas de l'écriture de A). L'espace $(E,\overline{\mathcal{B}},\mu)$ est alors complet et s'appelle le complété de $(E,\mathcal{B},\mu)$. Enfin on vérifie aisément que si $f : E \to \overline{\mathbb{R}}$ est $\overline{\mathcal{B}}$ mesurable, il existe $g, h : E \to \overline{\mathbb{R}}$, $\mathcal{B}$ mesurables telles que $g \leq f \leq h$ et $g = h$ p.p.

Dans la suite, la plupart du temps, on partira d'un espace mesurable ou d'un espace de probabilité sans se soucier de sa construction. Il est néanmoins indispensable de s'assurer de l'existence de tels objets. On va s'intéresser aux mesures sur $\mathcal{B}(\mathbb{R})$ finies sur les intervalles bornés. On verra une seconde méthode de construction en 1.5. Observons d'abord que $\mathcal{C} = \{\,]a,b],\ -\infty < a < b < +\infty\}$ est une classe stable par intersection finie et que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. Il résulte alors de la prop.1.2.3 qu'une mesure μ sur $\mathcal{B}(\mathbb{R})$ finie sur les intervalles bornés est déterminée par les valeurs $\mu(]a,b])$. Ensuite, étant donnée une telle mesure, si on pose

$$F(0) = 0;\ \ F(x) = \mu(]0,x]),\ x > 0;\ \ F(x) = -\mu(]x,0]),\ x < 0,$$

$F(x)$ est une fonction continue à droite et croissante et l'on a $\mu(]a,b]) = F(b) - F(a)$. On est donc ramené au problème suivant. Soit F une application de $\mathbb{R}$ dans $\mathbb{R}$ continue à droite et croissante, existe-t-il une mesure μ sur $\mathcal{B}(\mathbb{R})$ telle que $\mu(]a,b]) = F(b) - F(a)$? Il est facile de décrire l'algèbre $\mathcal{A}$ engendrée par $\mathcal{C}$, on a

$$\mathcal{A} = \{\, A = \cup_{k=1}^n]a_k, b_k],\ -\infty \leq a_1 < b_1 < a_2 < \ldots < b_{n-1} < a_n < b_n \leq +\infty\}$$

en convenant que, si $b_n = +\infty$, $]a_n, b_n] =]a_n, +\infty[$. On définit μ sur $\mathcal{A}$ par $\mu(A) = \sum_{k=1}^n F(b_k) - F(a_k)$ où $F(+\infty) = \lim_{x \to +\infty} F(x)$, $F(-\infty) = \lim_{x \to -\infty} F(x)$. Il est facile de montrer que μ est additive sur $\mathcal{A}$, un peu plus délicat de montrer que μ est σ-additive sur $\mathcal{A}$ mais cela se fait. On a donc construit une mesure μ sur $\mathcal{A}$ telle que $\mu(]a,b]) = F(b) - F(a)$. Pour passer à $\mathcal{B}(\mathbb{R})$, on utilise le théorème de Carathéodory:

Théorème 1.2.4. *Soit μ une mesure sur une algèbre $\mathcal{A}$, alors μ se prolonge en une mesure sur $\sigma(\mathcal{A})$. De plus, si μ est σ-finie, ce prolongement est unique.*

Tout ceci donne, puisque dans notre cas $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$,

Théorème 1.2.5. *Soit F une application de $\mathbb{R}$ dans $\mathbb{R}$ continue à droite et croissante. Il existe une et une seule mesure μ sur $\mathcal{B}(\mathbb{R})$ telle que, pour tous $a < b$, $\mu(]a,b]) = F(b) - F(a)$.*

Si on choisit $F(x) = x$, on obtient l'existence et l'unicité d'une mesure λ sur $\mathcal{B}(\mathbb{R})$ vérifiant, pour tout intervalle I, $\lambda(I) = |I|$. C'est la mesure de Lebesgue sur $\mathbb{R}$. Si $\mathcal{N}$ est la classe des ensembles λ-négligeables, $\overline{\mathcal{B}(\mathbb{R})} = \sigma(\mathcal{B}, \mathcal{N})$ s'appelle la tribu des ensembles Lebesgue-mesurables (elle est beaucoup plus "grosse" que $\mathcal{B}(\mathbb{R})$) et λ se prolonge sans peine à $\overline{\mathcal{B}(\mathbb{R})}$ comme en 1.2.3.

1.3. Intégration

Soit $(E, \mathcal{B}, \mu)$ un espace mesuré.

On va construire l'intégrale des fonctions positives par rapport à μ. Si $f \in e\mathcal{B}^+$, c'est très facile, f s'écrit $f = \sum_{k=1}^n a_k \mathbb{1}_{\{A_k\}}$, $A_k \in \mathcal{B}$ et l'on pose

$$\int f\, d\mu := \sum_{k=1}^n a_k \mu(A_k).$$

Des considérations élémentaires montrent que ceci ne dépend pas de l'écriture de f et que, pour $f, g \in \mathcal{B}_e^+$ et $a, b \in \mathbb{R}^+$, $\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu$ et que, si $f \le g$, $\int f\, d\mu \le \int g\, d\mu$. On a aussi le résultat plus technique suivant qui est la clé de la construction.

Lemme 1.3.1. *Si $f_n, g_n \in \mathcal{B}_e^+$ sont croissantes et si $\lim \uparrow f_n = \lim \uparrow g_n$, on a $\lim \uparrow \int f_n\, d\mu = \lim \uparrow \int g_n\, d\mu$.*

Soit $f \in \mathcal{B}^+$. Il existe (prop.1.1.2) une suite $f_n \in e\mathcal{B}^+$ telle que $f_n \uparrow f$, on a alors $\int f_n\, d\mu \uparrow$ et on pose $\int f\, d\mu = \lim \uparrow \int f_n\, d\mu$. Le point important est que, d'après le lem.1.3.1, cette limite ne dépend pas de la suite f_n choisie. On a en particulier, vu (1.2), pour $f \in \mathcal{B}^+$,

$$\int f\, d\mu = \lim \uparrow \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mu(\{x,\ \frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\}) + n\mu(\{x,\ f(x) \ge n\}).$$

Par passage à la limite, on obtient immédiatement que, pour $f, g \in \mathcal{B}^+$ et $a, b \in \mathbb{R}^+$, $\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu$ et que, si $f \le g$, $\int f\, d\mu \le \int g\, d\mu$. Enfin on dira que $f \in \mathcal{B}^+$ est intégrable si $\int f\, d\mu < +\infty$.

Pour l' intégration des fonctions réelles ou complexes on pose

$$\mathcal{L}^1 = \mathcal{L}^1(E, \mathcal{B}, \mu) = \{f \in \mathcal{B}_r,\ \int |f|\, d\mu < +\infty\}. \tag{1.3}$$

Si $f \in \mathcal{L}^1$, f^+ et f^- sont intégrables et on pose

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu.$$

Il est facile de voir (vu que $|f+g| \leq |f|+|g|$) que $\mathcal{L}^1$ est un espace vectoriel et que $f \mapsto \int f\,d\mu$ est une forme linéaire positive sur $\mathcal{L}^1$. De plus, pour $f \in \mathcal{L}^1$, $|\int f\,d\mu| \leq \int |f|\,d\mu$. Si f est $\mathcal{B}$-mesurable à valeurs $\mathbb{C}$, on pose ($|f|$ désignant le module),

$$\mathcal{L}^1_{\mathbb{C}} = \mathcal{L}^1_{\mathbb{C}}(E,\mathcal{B},\mu) = \{f\ \mathcal{B}\text{-mesurable complexe}, \int |f|\,d\mu < +\infty\}. \tag{1.4}$$

On définit alors, pour $f \in \mathcal{L}^1_{\mathbb{C}}$, $\int f\,d\mu = \int \Re(f)\,d\mu + i\int \Im(f)\,d\mu$. $\mathcal{L}^1_{\mathbb{C}}$ est un espace vectoriel sur $\mathbb{C}$ et $f \mapsto \int f\,d\mu$ une forme linéaire sur $\mathcal{L}^1_{\mathbb{C}}$. On a aussi, pour $f \in \mathcal{L}^1_{\mathbb{C}}$, $|\int f\,d\mu| \leq \int |f|\,d\mu$.

Propriétés.

- (i) Si $f \in \mathcal{B}^+$ et si $\int f\,d\mu < +\infty$, $f < +\infty$ p.p.
- (ii) Si $f \in \mathcal{B}^+$ et si $\int f\,d\mu = 0$, $f = 0$ p.p.
- (iii) Si $f \in \mathcal{L}^1_{\mathbb{C}}$ et si $A \in \mathcal{B}$, $f\mathbb{1}_{\{A\}} \in \mathcal{L}^1_{\mathbb{C}}$. On pose alors

$$\int_A f\,d\mu := \int f\mathbb{1}_{\{A\}}\,d\mu,\ \ A \in \mathcal{B},\ f \in \mathcal{L}^1_{\mathbb{C}} \cup \mathcal{B}^+.$$

- (iv) Si $f \in \mathcal{L}^1$ et si, pour tout $A \in \mathcal{B}$, $\int_A f\,d\mu \geq 0$ alors $f \geq 0$ p.p.

Il nous reste à énoncer les résultats concernant les passages à la limite. Le premier d'où découlent facilement les autres s'appelle théorème de convergence monotone ou théorème de Beppo-Levi.

Théorème 1.3.2. *Soit $f_n \in \mathcal{B}^+$ une suite croissante, alors*

$$\lim\uparrow \int f_n\,d\mu = \int \lim\uparrow f_n\,d\mu.$$

Corollaire 1.3.3. *Soit $g_n \in \mathcal{B}^+$, alors*

$$\sum_n \int g_n\,d\mu = \int \sum_n g_n\,d\mu.$$

Proposition 1.3.4. *(Lemme de Fatou)*

- *(i) Soit $f_n \in \mathcal{B}^+$, alors $\int \underline{\lim}\, f_n\,d\mu \leq \underline{\lim} \int f_n\,d\mu$.*
- *(ii) Soit $f_n \in \mathcal{B}_r$ avec $|f_n| \leq g \in \mathcal{L}^1$, alors*

$$\int \underline{\lim}\, f_n\,d\mu \leq \underline{\lim} \int f_n\,d\mu \leq \overline{\lim} \int f_n\,d\mu \leq \int \overline{\lim}\, f_n\,d\mu.$$

(ii) implique le célèbre théorème de Lebesgue,

Théorème 1.3.5. *Soit* $f_n \in \mathcal{L}^1_{\mathbb{C}}$ *telles que* $f_n \to f$ *p.p. avec* $|f_n| \le g \in \mathcal{L}^1$*, alors*

$$\lim \int f_n \, d\mu = \int f \, d\mu.$$

Ce théorème a une version "continue" très utile.

Corollaire 1.3.6. *Soit* $(f_t,\, t \in U)$ *une famille d'éléments de* $\mathcal{L}^1_{\mathbb{C}}$*,* U *ouvert de* $\mathbb{R}^d$*. On suppose que* $\lim_{t\to t_0} f_t = f$ *p.p. et que, pour tout* $t \in U$*,* $|f_t| \le g \in \mathcal{L}^1$*, alors* $\lim_{t\to t_0} \int f_t \, d\mu = \int f \, d\mu$*.*

`Preuve:`
Il suffit de remarquer que $\lim_{t\to t_0} \int f_t \, d\mu = \int f \, d\mu$ ssi, pour toute suite t_n tendant vers t_0, $\lim_{t_n\to t_0} \int f_{t_n} \, d\mu = \int f \, d\mu$ et d'appliquer le th.1.3.5 ∎

Donnons un exemple d'utilisation de ce corollaire.

Proposition 1.3.7. *Soient* $(E, \mathcal{B}, \mu)$ *un espace mesuré,* I *un intervalle ouvert de* $\mathbb{R}$ *et une famille* $(f(t,x),\, t \in I)$ *d'éléments de* $\mathcal{L}^1_{\mathbb{C}}(\mu)$*. On pose, pour tout* $t \in I$*,* $\phi(t) = \int f(t,x) \, d\mu(x)$*. On suppose que, pour tout* $x \in A$*,* $t \mapsto f(t,x)$ *est dérivable sur* I*, que, pour tous* $x \in A$ *et* $t \in I$*,* $|\frac{\partial f}{\partial t}(t,x)| \le g(x)$*, que* $g \in \mathcal{L}^1(\mu)$ *et que* $\mu(A^c) = 0$*. Alors* ϕ *est dérivable sur* I *et* $\phi'(t) = \int \frac{\partial f}{\partial t}(t,x) \, d\mu(x)$*.*

`Preuve:`
On a

$$\frac{1}{h}(\phi(t+h) - \phi(t)) = \int_A \frac{1}{h}(f(t+h,x) - f(t,x)) \, d\mu(x).$$

D'après la formule des accroissements finis, on a, pour $x \in A$,

$$|\frac{1}{h}(f(t+h,x) - f(t,x))| = |\frac{\partial f}{\partial t}(\theta,x)| \le g(x)$$

si h est assez petit et

$$\frac{1}{h}(f(t+h,x) - f(t,x)) \to_{h\to 0} \frac{\partial f}{\partial t}(t,x).$$

On peut appliquer le cor.1.3.6 et

$$\int_A \frac{1}{h}(f(t+h,x) - f(t,x)) \, d\mu(x) \to_{h\to 0} \int_A \frac{\partial f}{\partial t}(t,x) \, d\mu(x) = \int \frac{\partial f}{\partial t}(t,x) \, d\mu(x) \quad ∎$$

Lien avec l'intégrale usuelle. Soit f une fonction réelle continue sur $[a,b]$ et posons, pour $a \le x \le b$, $F(x) = \int_a^x f(t) \, dt$ (intégrale au sens usuelle) et $G(x) = \int \mathbb{1}_{\{[a,a+x[\}} f \, d\lambda$, λ mesure de Lebesgue sur $\mathbb{R}$. On sait que $F(a) = 0$, F est continue sur $[a,b]$ et que, sur $]a,b[$, F est dérivable avec $F' = f$. Il est facile de vérifier que G a les mêmes propriétés. Ceci implique que $F = G$ sur $[a,b]$ et, en particulier, que

$$\int_a^b f(t) \, dt = \int \mathbb{1}_{\{[a,b[\}} f \, d\lambda.$$

Par additivité, cette formule est encore vraie si f est continue par morceaux sur $[a,b]$. Considérons maintenant une application f de $\mathbb{R}$ dans $\mathbb{R}$ continue par morceaux telle que $\int_{-\infty}^{+\infty} f(t)\,dt$ soit absolument convergente. Lorsque $a \downarrow -\infty$ et $b \uparrow +\infty$, d'une part, par définition, $\int_a^b |f(t)|\,dt \to \int_{-\infty}^{+\infty} |f(t)|\,dt < +\infty$ et $\int_a^b f(t)\,dt \to \int_{-\infty}^{+\infty} f(t)\,dt$; d'autre part, $\int \mathbb{1}_{\{[a,b[\}}|f|\,d\lambda \to \int |f|\,d\lambda$ (convergence monotone) ce qui implique que $f \in \mathcal{L}^1(\lambda)$ puis $\int \mathbb{1}_{\{[a,b[\}} f\,d\lambda \to \int f\,d\lambda$ (théorème de Lebesgue puisque $|\mathbb{1}_{\{[a,b[\}} f| \leq |f| \in \mathcal{L}^1(\lambda)$). Donc

$$\int_{-\infty}^{+\infty} f(t)\,dt = \int f\,d\lambda.$$

Par contre, si $\int_{-\infty}^{+\infty} f(t)\,dt$ est convergente mais pas absolument convergente (par exemple $f(x) = \frac{\sin x}{x}$), $f \notin \mathcal{L}^1(\lambda)$.

Soient E un ensemble dénombrable et $(\mu(x),\, x \in E)$ une famille d'éléments de $\overline{\mathbb{R}}^+$. On pose, pour $A \subset E$, $\mu(A) = \sum_{x\in A} \mu(x)$. Le th.1.2.1 implique que μ est une mesure sur $(E, \mathcal{P}(E))$. On a alors $\mathcal{L}^1 = \{f,\, \sum_{x\in E} |f(x)|\mu(x) < +\infty\}$ et, pour $f \in \mathcal{L}^1$, $\int f\,d\mu = \sum_{x\in E} f(x)\mu(x)$. En particulier si on prend pour μ la mesure de comptage i.e. $\mu(x) = 1$ pour tout $x \in E$, on a $\mathcal{L}^1 = \{f,\, \sum_{x\in E} |f(x)| < +\infty\}$ et $\int f\,d\mu = \sum_{x\in E} f(x)$. Il est intéressant d'énoncer dans ce cadre les théorèmes de convergence du paragraphe précédent. On a

- (i) Si $0 \leq f_n \uparrow f$, $\sum_x f_n(x) \uparrow \sum_x f(x)$.
- (ii) Si $0 \leq f_n$, $\sum_x \underline{\lim}_n f_n(x) \leq \underline{\lim}_n \sum_x f_n(x)$.
- (iii) Si $f_n \to f$ et si $|f_n| \leq g$ avec $\sum_x g(x) < +\infty$, $\sum_x f_n(x) \to_n \sum_x f(x)$.

/bf Espaces L^p.

Soit $(E, \mathcal{B}, \mu)$ un espace mesuré. On note $\mathcal{L}^0$ l'ensemble des applications $\mathcal{B}$-mesurables de E dans $\overline{\mathbb{R}}$ finies p.p. On dit que $f \sim g$ si $f = g$ p.p. Alors $\sim$ est une relation d'équivalence sur $\mathcal{L}^0$. On note $L^0 = \mathcal{L}^0/\sim$. En fait L^0 est l'espace des classes de fonctions $\mathcal{B}$-mesurables définies à un p.p. près. Puisque $f = g$ p.p. implique $\int |f|\,d\mu = \int |g|\,d\mu$ et $\int f\,d\mu = \int g\,d\mu$ si f et g sont dans $\mathcal{L}^1$, on peut définir sans ambiguïté, pour $f \in L^0$, $\int |f|\,d\mu$ puis, si $\int |f|\,d\mu < +\infty$, $\int f\,d\mu$. Par abus de langage, *dans toute la suite nous noterons de la même façon une fonction et sa classe d'équivalence.* On pose alors, pour $1 \leq p < +\infty$ et $f \in L^0$,

$$||f||_p = [\int |f|^p\,d\mu]^{\frac{1}{p}}$$

et, pour $p = +\infty$,

$$||f||_\infty = \inf(M,\, \mu(|f| > M) = 0).$$

On a deux inégalités fondamentales. Pour $f, g \in L^0_+$,

$$||f+g||_p \leq ||f||_p + ||g||_p, \;\; 1 \leq p \leq +\infty \tag{1.5}$$

qui s'appelle l'inégalité de Minkowski et

$$||fg||_1 \leq ||f||_p \, ||g||_q, \;\; 1 \leq p \leq +\infty, \; \frac{1}{p} + \frac{1}{q} = 1 \tag{1.6}$$

qui s'appelle l'inégalité de Hölder. Notons que pour $p = q = 2$, (1.6) implique l'inégalité de Schwarz

$$(\int |fg| \, d\mu)^2 \leq (\int f^2 \, d\mu)(\int g^2 \, d\mu).$$

On note

$$\mathcal{L}^p = \{f \in \mathcal{L}^0, \; \int |f|^p \, d\mu < +\infty\}, \quad L^p = \{f \in L^0, \; \int |f|^p \, d\mu < +\infty\}.$$

Alors L^p muni de la norme $||.||_p$ est un espace de Banach et L^2 est un espace de Hilbert pour le produit scalaire

$$< f, g >= \int fg \, d\mu.$$

On peut aussi considérer le cas des fonctions à valeurs complexes. On définit de la même façon $L^p_{\mathbb{C}} = L^p_{\mathbb{C}}(E, \mathcal{B}, \mu)$. Il faut noter que $L^2_{\mathbb{C}}$ est associé au produit scalaire

$$< f, g >= \int f\bar{g} \, d\mu.$$

Proposition 1.3.8. *Pour* $1 \leq p < +\infty$, $\mathcal{E}^0 = \{f, \; f = \sum_{k=1}^{n} a_k \mathbb{1}_{\{A_k\}}, \; A_k \in \mathcal{B}, \; \mu(A_k) < +\infty\}$ *est dense dans* $L^p(E, \mathcal{B}, \mu)$.

`Preuve:`
Il suffit de considérer $f \geq 0$. Alors il existe (prop.1.1.2) une suite $f_n \in \mathcal{B}^+_e$ telle que $f_n \uparrow f$. Vu que $f_n^p \leq f^p \in \mathcal{L}^1$, $f_n \in \mathcal{E}^0$. On a, puisque $f < +\infty$ p.p., $|f - f_n|^p \to 0$ p.p. et $|f - f_n|^p \leq f^p \in \mathcal{L}^1$ donc (th. de Lebesgue) $\int |f - f_n|^p \, d\mu \to 0$ ∎

Soit μ une mesure sur $(E, \mathcal{B})$. On peut lui associer une application I de $\mathcal{B}^+$ dans $\overline{\mathbb{R}}^+$ en posant $I(f) = \int f \, d\mu$, $f \in \mathcal{B}^+$. L'application I a les propriétés suivantes: $I(f+g) = I(f) + I(g)$, $I(af) = aI(f)$, $a \in \mathbb{R}^+$ et $I(f_n) \uparrow I(f)$ si $f_n \uparrow f$. Réciproquement on a,

Proposition 1.3.9. *Soient* $(E, \mathcal{B})$ *un espace mesurable et* I *une application de* $\mathcal{B}^+$ *dans* $\overline{\mathbb{R}}^+$ *telle que*

- *(i) si* $f, g \in \mathcal{B}^+$, $I(f+g) = I(f) + I(g)$; *si* $f \in \mathcal{B}^+$ *et* $a \in \mathbb{R}^+$, $I(af) = aI(f)$,
- *(ii) si* $f_n \in \mathcal{B}^+$ *et si* $f_n \uparrow f$, $I(f_n) \uparrow I(f)$.

Alors $\mu(A) = I(\mathbb{1}_{\{A\}})$, $A \in \mathcal{B}$, *définit une mesure sur* $\mathcal{B}$ *et on a, pour toute* $f \in \mathcal{B}^+$, $I(f) = \int f \, d\mu$.

`Preuve:`
Soient $A_n \in \mathcal{B}$ des ensembles deux à deux disjoints d'union A, on a $\mathbb{1}_{\{A\}} = \sum_n \mathbb{1}_{\{A_n\}} = \lim \uparrow \sum_{k=1}^n \mathbb{1}_{\{A_k\}}$ et

$$\mu(A) = I(\mathbb{1}_{\{A\}}) = I(\lim \uparrow \sum_{k=1}^{n} \mathbb{1}_{\{A_k\}}) = \lim \uparrow I(\sum_{k=1}^{n} \mathbb{1}_{\{A_k\}}) = \lim \uparrow \sum_{k=1}^{n} I(\mathbb{1}_{\{A_k\}}) = \sum_n \mu(A_n).$$

Ce qui montre que μ est une mesure. On a alors, pour toute $f \in e\mathcal{B}^+$, $I(f) = \int f\,d\mu$. On conclut facilement en utilisant la prop.1.1.2 ∎

Donnons deux applications.

Mesures à densité.

Proposition 1.3.10. *Soient $(E, \mathcal{B}, \mu)$ un espace mesuré et $h \in \mathcal{B}^+$. La formule $\nu(A) = \int_A h\,d\mu$, $A \in \mathcal{B}$ définit une mesure sur $\mathcal{B}$ appelée mesure de densité h par rapport à μ. On a, pour toute $f \in \mathcal{B}^+$,*

$$\int f\,d\nu = \int fh\,d\mu. \tag{1.7}$$

De plus $f \in \mathcal{B}_r$ est ν-intégrable ssi fh est μ-intégrable et l'on a dans ce cas (1.7).

`Preuve:`
On considère la fonctionnelle $I(f) = \int fh\,d\mu$, $f \in \mathcal{B}^+$ et on applique la prop.1.3.9. La dernière assertion est pure routine en écrivant $f = f^+ - f^-$ ∎

Mesures images.

Proposition 1.3.11. *Soient h une application mesurable de $(E, \mathcal{B})$ dans $(F, \mathcal{F})$ et μ une mesure sur $(E, \mathcal{B})$. La formule $\nu(A) = \mu(h^{-1}(A))$, $A \in \mathcal{F}$, définit une mesure sur $(F, \mathcal{F})$ appelée mesure image de μ par h. On a, pour toute $f \in \mathcal{F}^+$,*

$$\int f\,d\nu = \int f \circ h\,d\mu. \tag{1.8}$$

De plus $f \in \mathcal{F}_r$ est ν-intégrable ssi $f \circ h$ est μ-intégrable et l'on a dans ce cas (1.8).

`Preuve:`
On considère la fonctionnelle $I(f) = \int f \circ h\,d\mu$, $f \in \mathcal{F}^+$ et on applique la prop.1.3.9. La mesure associée à I est

$$\nu(A) = I(A) = \int \mathbb{1}_{\{A\}} \circ h\,d\mu = \int \mathbb{1}_{\{h^{-1}(A)\}}\,d\mu = \mu(h^{-1}(A)).$$

On conclut facilement ∎

1.4. Mesures produits

Soient $(E_1,\mathcal{B}_1)$ $(E_2,\mathcal{B}_2)$ deux espaces mesurables. On définit une tribu sur $E_1 \times E_2$, appelée tribu produit de $\mathcal{B}_1$ et $\mathcal{B}_2$ et notée $\mathcal{B}_1 \otimes \mathcal{B}_2$, par

$$\mathcal{B}_1 \otimes \mathcal{B}_2 = \sigma(A_1 \times A_2,\ A_1 \in \mathcal{B}_1,\ A_2 \in \mathcal{B}_2).$$

Alors si $f : E_1 \times E_2 \to \overline{\mathbb{R}}^+$ est une fonction $\mathcal{B}_1 \otimes \mathcal{B}_2$-mesurable, on a que pour tout $x_1 \in E_1$, $x_2 \mapsto f(x_1,x_2)$ est $\mathcal{B}_2$-mesurable et que, pour tout $x_2 \in E_2$, $x_1 \mapsto f(x_1,x_2)$ est $\mathcal{B}_1$-mesurable. En particulier si $A \in \mathcal{B}_1 \otimes \mathcal{B}_2$, $A_{x_2} = \{x_1, (x_1,x_2) \in A\} \in \mathcal{B}_1$ et $A_{x_1} = \{x_2, (x_1,x_2) \in A\} \in \mathcal{B}_2$. On en déduit facilement que, si $f \in (\mathcal{B}_1 \otimes \mathcal{B}_2)^+$ et si μ_i est une mesure sur $(E_i,\mathcal{B}_i)$, $x_1 \mapsto \int f(x_1,x_2)\,d\mu_2(x_2)$ est $\mathcal{B}_1$-mesurable et $x_2 \mapsto \int f(x_1,x_2)\,d\mu_1(x_1)$ est $\mathcal{B}_2$-mesurable.

Théorème 1.4.1. *Soient $(E_1,\mathcal{B}_1,\mu_1)$ et $(E_2,\mathcal{B}_2,\mu_2)$ deux espaces mesurés avec μ_1 et μ_2 σ-finies. Il existe une unique mesure sur $\mathcal{B}_1 \otimes \mathcal{B}_2$, notée $\mu_1 \otimes \mu_2$ et appelée mesure produit de μ_1 et μ_2, telle que,*

$$\textit{pour tous } A_1 \in \mathcal{B}_1,\ A_2 \in \mathcal{B}_2,\ \ \mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\,\mu(A_2).$$

De plus, pour toute $f \in (\mathcal{B}_1 \otimes \mathcal{B}_2)^+$,

$$\int f\,d\mu_1 \otimes \mu_2 = \int (\int f(x_1,x_2)\,d\mu_1(x_1))\,d\mu_2(x_2) = \int (\int f(x_1,x_2)\,d\mu_2(x_2))\,d\mu_1(x_1).$$

`Preuve:`
(i) Unicité. On applique la prop.1.2.3 à $\mathcal{C} = \{A,\ A = A_1 \times A_2,\ A_1 \in \mathcal{B}_1,\ A_2 \in \mathcal{B}_2,\ \mu(A_1) < +\infty,\ \mu(A_2) < +\infty\}$.

(ii) Existence. On applique la prop.1.3.9 à $I_1(f) = \int(\int f(x_1,x_2)\,d\mu_1(x_1))\,d\mu_2(x_2)$ ce qui donne l'existence. Mais on peut aussi appliquer la prop.1.3.9 à $I_2(f) = \int(\int f(x_1,x_2)\,d\mu_2(x_2))\,d\mu_1(x_1)$ et, vu l'unicité, on a $I_1(f) = I_2(f)$ ∎

Si $f \in \mathcal{L}^1_{\mathbb{C}}(\mu_1 \otimes \mu_2)$, on peut appliquer le théorème précédent à $(\Re(f))^+$, $(\Re(f))^-$, $(\Im(f))^+$ et $(\Im(f))^-$ et l'on obtient le théorème de Fubini:

Théorème 1.4.2. *Soit $f \in \mathcal{L}^1_{\mathbb{C}}(\mu_1 \otimes \mu_2)$. Alors,*
$\int |f(x_1,x_2)|\,d\mu_2(x_2) < +\infty$ μ_1 p.p. et $\phi_1(x_1) = \int f(x_1,x_2)\,d\mu_2(x_2) \in \mathcal{L}^1(\mu_1)$,
$\int |f(x_1,x_2)|\,d\mu_1(x_1) < +\infty$ μ_2 p.p. et $\phi_2(x_2) = \int f(x_1,x_2)\,d\mu_1(x_1) \in \mathcal{L}^1(\mu_2)$, et

$$\int f\,d\mu_1 \otimes \mu_2 = \int (\int f(x_1,x_2)\,d\mu_1(x_1))\,d\mu_2(x_2) = \int (\int f(x_1,x_2)\,d\mu_2(x_2))\,d\mu_1(x_1).$$

Tout ceci s'étend sans (trop de) peine au cas de n espaces mesurables. Il y a quelques vérifications fastidieuses à faire du type $\mu_1 \otimes (\mu_2 \otimes \mu_3) = (\mu_1 \otimes \mu_2) \otimes \mu_3$. De plus dans la formule d'intégrations successives, les variables peuvent être intégrées dans tous les ordres possibles. A ce sujet, le grand principe est: si f est positive, tout est permis, si f est de signe quelconque ou complexe, on considère d'abord $|f|$ et on commence par montrer que $|f|$ est intégrable.

Considérons $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, λ mesure de Lebesgue. La première chose à vérifier est que $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$, ce qui est assez facile. On définit alors, sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\lambda_d = \lambda \otimes \lambda \otimes \ldots \otimes \lambda$. On peut appliquer la prop.1.2.3 à

$$\mathcal{C} = \{A,\ A = \prod_{i=1}^{d}]\, a_i, b_i[,\ -\infty < a_i < b_i < +\infty\}.$$

On obtient que λ_d est l'unique mesure sur $\mathcal{B}(\mathbb{R}^d)$ telle que, pour tout $-\infty < a_i < b_i < +\infty$,

$$\lambda_d(\prod_{i=1}^{d}]a_i, b_i[) = \prod_{i=1}^{d} (b_i - a_i).$$

On appelle λ_d la mesure de Lebesgue de $\mathbb{R}^d$.

1.5. Mesures de Radon sur $\mathbb{R}^d$

Nous étudions plus en détails les mesures sur $\mathbb{R}^d$ utilisant ses propriétés topologiques.

On note:On note

- C_b l'ensemble des fonctions continues bornées sur $\mathbb{R}^d$,
- C_0 l'ensemble des fonctions de C_b tendant vers 0 à l'infini,
- C_k l'ensemble des fonctions de C_0 à support compact.

On munit ces espaces de la norme $||f|| = \sup_{x \in \mathbb{R}^d} |f(x)|$. Alors C_0 est un espace de Banach séparable (il existe une suite dense). Rappelons que:

- Etant donnés K compact $\subset U$ ouvert, il existe $f \in C_k$, $0 \leq f \leq 1$, $f = 1$ sur K, $f = 0$ sur U^c.
- Soit K un compact. Il existe une suite d'ouverts U_n et une suite $f_n \in C_k$ telles que $K = \lim \downarrow U_n$, $\mathbb{1}_{\{K\}} = \lim \downarrow f_n$.
- Soit U un ouvert. Il existe une suite de compacts K_n et une suite $f_n \in C_k$ telles que $U = \lim \uparrow K_n$, $\mathbb{1}_{\{U\}} = \lim \uparrow f_n$.

L'objet auquel on s'intéresse est le suivant.

Définition 1.5.1. *On appelle mesure de Radon sur $\mathbb{R}^d$ toute mesure sur $\mathcal{B}(\mathbb{R}^d)$ finie sur tout compact.*

Donc toute mesure bornée sur $\mathcal{B}(\mathbb{R}^d)$ ainsi que la mesure de Lebesgue sur $\mathbb{R}^d$ sont des mesures de Radon.
Soit μ une mesure de Radon sur $\mathbb{R}^d$. Alors $C_k \subset \mathcal{L}^1(\mu)$ et $I : f \mapsto \int f\, d\mu$ est une forme linéaire positive sur C_k. Il est remarquable que toutes les mesures de Radon s'obtiennent ainsi (c'est le théorème de Riesz):

Théorème 1.5.2. *Soit I une forme linéaire positive sur C_k. Il existe une et une seule mesure de Radon μ sur $\mathcal{B}(\mathbb{R}^d)$ telle que, pour toute $f \in C_k$, $I(f) = \int f\,d\mu$.*

Ce théorème fournit une seconde méthode pour construire des mesures. Considérons par exemple $C_k(\mathbb{R})$. Pour $f \in C_k(\mathbb{R})$, on peut construire $\int_{-\infty}^{+\infty} f(x)\,dx$ par des méthodes élémentaires (sommes de Riemann ou limite de fonctions en escalier). Appliquant ce théorème, on en déduit l'existence de la mesure de Lebesgue sur $\mathbb{R}$. La démonstration du th.1.5.2 est difficile. Notons cependant que l'unicité est une conséquence immédiate de la prop.1.2.3. En effet:

Lemme 1.5.3. *Soient μ_1 et μ_2 des mesures de Radon sur $\mathcal{B}(\mathbb{R}^d)$. On suppose que, pour toute $f \in C_k$, $\int f\,d\mu_1 = \int f\,d\mu_2$. Alors $\mu_1 = \mu_2$.*

`Preuve:`
Soit $\mathcal{C}$ la classe des ouverts bornés. Pour tout $U \in \mathcal{C}$, il existe $f_n \in C_k$ telle que $\mathbb{1}_{\{U\}} = \lim \uparrow f_n$. On a donc (convergence monotone) $\mu_1(U) = \mu_2(U) < +\infty$. Comme $\mathcal{C}$ est stable par intersection finie, engendre $\mathcal{B}(\mathbb{R}^d)$ et que $\mathbb{R}^d = \lim \uparrow U_n$, $U_n \in \mathcal{C}$, on conclut par la prop.1.2.3 ∎

Si la mesure μ est bornée, $C_0 \subset \mathcal{L}^1(\mu)$ et $I : f \mapsto \int f\,d\mu$ est une forme linéaire positive sur C_0. On a réciproquement:

Théorème 1.5.4. *Soit I une forme linéaire positive sur C_0. Il existe une et une seule mesure bornée μ sur $\mathcal{B}(\mathbb{R}^d)$ telle que, pour toute $f \in C_0$, $I(f) = \int f\,d\mu$.*

Toute mesure de Radon est régulière au sens suivant.

Proposition 1.5.5. *Soit μ une mesure de Radon sur $\mathcal{B}(\mathbb{R}^d)$. On a, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mu(A) = \sup(\mu(K),\ K \text{ compact} \subset A) = \inf(\mu(U),\ U \text{ ouvert} \supset A).$$

`Preuve:`
(i) On suppose d'abord μ bornée. Soit $\mathcal{C} = \{A \in \mathcal{B}(\mathbb{R}^d)$, pour tout $\varepsilon > 0$, il existe K compact et U ouvert tels que $K \subset A \subset U$ et $\mu(U \setminus K) < \varepsilon\}$. Alors

a. $\mathcal{C}$ contient les ouverts. Si U est un ouvert, il existe K_n compacts, $K_n \uparrow U$ et donc $\mu(U \setminus K_n) \to 0$.

b. $\mathcal{C}$ est stable par complémentation. Supposons $K \subset A \subset U$ et $\mu(U \setminus K) < \varepsilon$, on a $U^c \subset A^c \subset K^c$ et $\mu(K^c \setminus U^c) < \varepsilon$, K^c est ouvert et U^c fermé. On choisit K_n compacts, $K_n \uparrow \mathbb{R}^d$, $\mu(U^c \setminus U^c \cap K_n) \to 0$ et $U^c \cap K_n$ est un compact inclus dans A^c.

c. $\mathcal{C}$ est stable par intersection dénombrable. Soient $A_p \in \mathcal{C}$, $p \geq 1$. Il existe K_p compacts et U_p ouverts tels que $K_p \subset A_p \subset U_p$ et $\mu(U_p \setminus K_p) < \frac{\varepsilon}{2^p}$. On a $\cap K_p \subset \cap A_p \subset \cap U_p$ et $\mu(\cap U_p \setminus \cap K_p) \leq \mu(\cup(U_p \setminus K_p)) \leq \sum \mu(U_p \setminus K_p) < \varepsilon$. $K = \cap K_p$ est compact. Soit $V_n = \cap_{p=1}^n U_p$, V_n est ouvert, $V_n \supset \cap U_p$ et $\mu(V_n \setminus \cap U_p) \to 0$. Donc $\cap A_p \in \mathcal{C}$.

Les points a,b,c montrent que $\mathcal{C}$ est une tribu et donc $\mathcal{C} = \mathcal{B}(\mathbb{R}^d)$.

(ii) Passons à μ quelconque. Si $\mu(A) < +\infty$, on applique le résultat précédent à la mesure bornée $\nu(B) = \mu(A \cap B)$ ce qui donne l'approximation par des compacts. Pour les ouverts, on considère $V_n = \{|x| < n\}$, $D_n = V_n \setminus V_{n-1}$ et les mesures bornées $\mu_n(B) = \mu(V_n \cap B)$. Pour tous n et $\varepsilon > 0$, il existe vu (i) des ouverts U_n contenant $A \cap D_n$ tels que $\mu_n(U_n \cap V_n) < \mu(A \cap D_n) + \varepsilon 2^{-n}$. On a alors $A \subset U = \cup_n U_n \cap V_n$ ouvert et $\mu(U) \leq \sum_n \mu(U_n \cap V_n) < \mu(A) + \varepsilon$.
Si $\mu(A) = +\infty$, il n'y a rien à montrer pour les ouverts. On choisit alors K_n compacts, $K_n \uparrow \mathbb{R}^d$, on a $\mu(A \cap K_n) \uparrow \mu(A) = +\infty$ et, pour tout n, $\mu(A \cap K_n) = \sup\{\mu(H),\ H$ compact $\subset A \cap K_n\}$. On conclut facilement ∎

De cette proposition, on déduit aisément,

Théorème 1.5.6. *Soit μ une mesure de Radon sur $\mathcal{B}(\mathbb{R}^d)$. Alors pour tout p, $1 \leq p < +\infty$, C_k est dense dans $L^p(\mathbb{R}^d, \mu)$.*

`Preuve:`
Vu la prop.1.3.8, il suffit d'approcher $\mathbb{1}_{\{A\}}$, $A \in \mathcal{B}(\mathbb{R}^d)$, $\mu(A) < +\infty$. D'après la prop.1.5.5, il existe des compacts K_n tels que $K_n \subset A$ et $\mu(A \setminus K_n) = ||\mathbb{1}_{\{A\}} - \mathbb{1}_{\{K_n\}}||_p^p \to 0$. Il suffit donc d'approcher $\mathbb{1}_{\{K\}}$, K compact. Mais il existe $f_n \in C_k$ tels que $f_n \downarrow \mathbb{1}_{\{K\}}$. Utilisant le th. de Lebesgue, on a $||\mathbb{1}_{\{K\}} - f_n||_p \to_n 0$ ∎

Corollaire 1.5.7. *Soit μ une mesure de Radon sur $\mathcal{B}(\mathbb{R})$. L'espace des fonctions en escaliers i.e. de la forme $\sum_{k=1}^n a_i \mathbb{1}_{\{[t_i, t_{i+1}[\}}$, $t_1 < \ldots < t_n$, est dense dans $L^p(\mathbb{R}, \mu)$.*

`Preuve:`
Vu l'uniforme continuité, il est facile d'approcher en norme $||.||_p$ une fonction de C_k par des fonctions en escaliers. On applique ensuite la prop.1.3.8 ∎

1.6. Convolution et transformation de Fourier

Précisons d'abord que nous n'avons pas l'intention de faire un exposé un tant soit peu exhaustif du sujet mais seulement de présenter quelques résultats utiles en probabilités. On se place sur $\mathbb{R}^d$. On note $\mathcal{M}_b$ l'ensemble des mesures bornées sur $\mathcal{B}(\mathbb{R}^d)$.. Si $f \in C_b$ et $\mu \in \mathcal{M}_b$, on écrit indifféremment $\int f\, d\mu$, $\mu(f)$ ou $< \mu, f >$. On note λ la mesure de Lebesgue sur $\mathbb{R}^d$ et dx pour $d\lambda(x)$. Si $\mu \in \mathcal{M}_b$ a une densité h par rapport à λ, on écrit $\mu = h.\lambda$.

convolution

Soit ϕ l'application de $\mathbb{R}^d \times \mathbb{R}^d$ dans $\mathbb{R}^d$, $(x, y) \mapsto x + y$.

Définition 1.6.1. *Soient $\mu, \nu \in \mathcal{M}_b$. On appelle produit de convolution de μ et ν et on note $\mu * \nu$ la mesure sur $\mathbb{R}^d$ image de $\mu \otimes \nu$ par ϕ.*

D'après (1.8), $\mu * \nu$ est caractérisée par:

$$\text{pour toute } f \in \mathcal{B}^+(\mathbb{R}^d),\ \int f\, d(\mu * \nu) = \int f(x + y)\, d\mu(x) d\nu(y). \tag{1.9}$$

On a donc $\mu * \nu(1) = \mu(1)\nu(1)$ et $\mu * \nu \in \mathcal{M}_b$.

Supposons que $\mu = \phi.\lambda$. On a (tout est positif)

$$\int f\, d(\mu * \nu) = \int f(x+y)\phi(x)\, dx d\nu(y) = \int f(x)(\int \phi(x-y)\, d\nu(y))\, dx$$

et donc, si on pose

$$\phi * \nu(x) = \int \phi(x-y)\, d\nu(y), \tag{1.10}$$

$\phi * \nu < +\infty$ λ p.p. et on a $(\phi.\lambda) * \nu = \phi * \nu.\lambda$. Noter que (1.10) est définie sans problème pour $\phi \in \mathcal{L}^\infty(\nu)$ et $\nu \in \mathcal{M}_b$.

Supposons que $\mu = \phi.\lambda$ et $\nu = \psi.\lambda$. On a (tout est positif)

$$\int f\, d(\mu * \nu) = \int f(x+y)\phi(x)\psi(y)\, dx dy = \int f(x)(\int \phi(x-y)\psi(y)\, dy)\, dx,$$

et donc, si on pose

$$\phi * \psi(x) = \int \phi(x-y)\psi(y)\, dy, \tag{1.11}$$

$\phi * \psi < +\infty$ λ p.p. et on a $(\phi.\lambda) * (\psi.\lambda) = \phi * \psi.\lambda$. Noter que (1.11) est définie sans problème pour $\phi \in \mathcal{L}^\infty(\lambda)$ et $\psi \in \mathcal{L}^1(\lambda)$.

Transformation de Fourier

Définition 1.6.2. *Soit $\mu \in \mathcal{M}_b$. On appelle transformée de Fourier de μ et on note $\hat{\mu}$ la fonction sur $\mathbb{R}^d$ définie par $\hat{\mu}(t) = \int e^{i<t,x>}\, d\mu(x)$.*

Vu que $|e^{i<t,x>}| \leq 1 \in \mathcal{L}^1(\mu)$, $t \mapsto \hat{\mu}(t)$ est continue (cor.1.3.6). Si μ est symétrique (i.e. $\mu(A) = \mu(-A)$), $\hat{\mu}$ est réelle. Enfin on a

$$|\hat{\mu}(t)| \leq \mu(1) = \hat{\mu}(0). \tag{1.12}$$

Si on note

$$\hat{f}(t) = \int e^{i<t,x>} f(x)\, dx, \quad f \in \mathcal{L}^1(\lambda),$$

on a, pour $\mu = h.\lambda$, $\hat{\mu}(t) = \hat{h}(t)$.

Théorème 1.6.3. *Soient $\mu, \nu \in \mathcal{M}_b$. On a $\widehat{\mu * \nu}(t) = \hat{\mu}(t)\hat{\nu}(t)$.*

`Preuve:`
En effet, puisque $|e^{i<t,x+y>}| \leq 1 \in \mathcal{L}^1(\mu \otimes \nu)$, on a (th.1.4.2),

$$\widehat{\mu * \nu}(t) = \int e^{i<t,x>}\, d(\mu * \nu)(x) = \int e^{i<t,x+y>}\, d\mu(x) d\nu(y)$$
$$= \int e^{i<t,x>}\, d\mu(x) \int e^{i<t,y>}\, d\nu(x) = \hat{\mu}(t)\hat{\nu}(t) \blacksquare$$

Ce paragraphe est consacré à quelques résultats techniques utiles pour la suite.
On pose, pour $\sigma > 0$ et $x \in \mathbb{R}^d$,

$$g_\sigma(x) = (2\pi\sigma^2)^{-d/2} \exp(-\frac{|x|^2}{2\sigma^2}), \quad |x|^2 = x_1^2 + \ldots + x_d^2 . \tag{1.13}$$

On a $g_\sigma \in C_0$, $\int g_\sigma(x)\,dx = 1$, $\int |x|^2 g_\sigma(x)\,dx = \sigma^2 d$.

Lemme 1.6.4. $\widehat{g_\sigma}(t) = \exp(-\frac{\sigma^2|t|^2}{2})$.

`Preuve:`
Soit $\phi(t) = (2\pi)^{-1/2} \int \mathrm{e}^{itu}\mathrm{e}^{-u^2/2}\,du$, $t \in \mathbb{R}$. Vu que $|\frac{d}{dt}\mathrm{e}^{itu}| \le |u| \in \mathcal{L}^1(\mathrm{e}^{-u^2/2}.\lambda)$, on peut appliquer la prop.1.3.7 et on a

$$\phi'(t) = i(2\pi)^{-1/2} \int \mathrm{e}^{itu}\,d(-\mathrm{e}^{-u^2/2}) = -(2\pi)^{-1/2} t \int \mathrm{e}^{itu}\mathrm{e}^{-u^2/2}\,du = -t\phi(t)$$

d'où $\phi(t) = C\mathrm{e}^{-t^2/2} = \mathrm{e}^{-t^2/2}$ puisque $\phi(0) = 1$. Alors (th.1.4.2)

$$(2\pi\sigma^2)^{-d/2} \int \mathrm{e}^{i<t,x>}\mathrm{e}^{-|x|^2/2\sigma^2}\,dx = \prod_{k=1}^{d} (2\pi\sigma^2)^{-1/2} \int \mathrm{e}^{it_k x_k}\mathrm{e}^{-x_k^2/2\sigma^2}\,dx_k = \mathrm{e}^{-\sigma^2|t|^2/2} \quad \blacksquare$$

Lemme 1.6.5. *Pour toute $f \in C_0$, $||g_\sigma * f - f|| \to_{\sigma\to 0} 0$.*

`Preuve:`
Notons d'abord que

$$\alpha^2 \int_{\{|x|\ge\alpha\}} g_\sigma(x)\,dx \le \int_{\{|x|\ge\alpha\}} |x|^2 g_\sigma(x)\,dx \le \sigma^2 d.$$

Soit $\varepsilon > 0$. f étant uniformément continue, il existe $\alpha > 0$ tel que, pour tous x, y vérifiant $|x - y| < \alpha$, $|f(x) - f(y)| < \varepsilon/2$. Alors

$$|g_\sigma * f(x) - f(x)| = |\int g_\sigma(x-y)(f(x) - f(y))\,dy|$$

$$\le \int_{\{|x-y|<\alpha\}} g_\sigma(x-y)|f(x) - f(y)|\,dy + \int_{\{|x-y|\ge\alpha\}} g_\sigma(x-y)|f(x) - f(y)|\,dy$$

$$\le \frac{\varepsilon}{2} \int g_\sigma(x-y)\,dy + 2\,||f|| \int_{\{|x|\ge\alpha\}} g_\sigma(x)\,dx \le \frac{\varepsilon}{2} + 2||f||\frac{\sigma^2 d}{\alpha^2} \le \varepsilon$$

si σ est assez petit $\blacksquare$

Corollaire 1.6.6. *La famille $(g_\sigma * f;\ \sigma > 0,\ f \in C_k)$ est dense dans C_0.*

`Preuve:`
On a $g_\sigma * f(x) = \int g_\sigma(x-y)f(y)\,dy$. Vu que $|g_\sigma(x-y)f(y))| \le |f(y)| \in \mathcal{L}^1(\lambda)$, on montre facilement, grâce au cor.1.3.6, que $g_\sigma * f \in C_0$. C_k étant dense dans C_0, il suffit d'approcher les fonctions de C_k, c'est l'objet du lem.1.6.5 $\blacksquare$

Lemme 1.6.7. *Pour toutes $\mu, \nu \in \mathcal{M}_b$,*

$$\int g_\sigma * \nu(x)\, d\mu(x) = (2\pi)^{-d/2} \int e^{-i<y,u>} g_1(\sigma u)\hat{\mu}(u)\, dud\nu(y).$$

`Preuve:`
Vu que, $g_\sigma(x) = \sigma^{-d} g_1(x/\sigma)$, on a (lem.1.6.4),

$$\mathrm{e}^{-|t|^2/2\sigma^2} = \hat{g}_{\sigma^{-1}}(t) = \int \mathrm{e}^{i<t,u>} g_{\sigma^{-1}}(u)\, du = \sigma^d \int \mathrm{e}^{i<t,u>} g_1(\sigma u)\, du$$

et $g_\sigma(t) = (2\pi)^{-d/2} \int \mathrm{e}^{i<t,u>} g_1(\sigma u)\, du$. On a alors

$$\begin{aligned} \int g_\sigma * \nu(x)\, d\mu(x) &= \int g_\sigma(x-y)\, d\nu(y) d\mu(x) \\ &= (2\pi)^{-d/2} \int \mathrm{e}^{i<x-y,u>} g_1(\sigma u)\, dud\mu(x)d\nu(y) \\ &= (2\pi)^{-d/2} \int \mathrm{e}^{-i<y,u>} g_1(\sigma u) (\int \mathrm{e}^{i<x,u>}\, d\mu(x))\, dud\nu(y) \end{aligned}$$

grâce au th.1.4.2 puisque $|\mathrm{e}^{i<x-y,u>} g_1(\sigma u)| \le |g_1(\sigma u)| \in \mathcal{L}^1(\lambda \otimes \mu \otimes \nu)$ ∎

Prenant $\nu = \delta_a$ dans le lem.1.6.7, on a, pour toute $\mu \in \mathcal{M}_b$,

$$\int g_\sigma(x-a)\, d\mu(x) = (2\pi)^{-d/2} \int \mathrm{e}^{-i<a,u>} g_1(\sigma u)\hat{\mu}(u)\, du. \tag{1.14}$$

Si on prend $\nu = f.\lambda$ dans le lem.1.6.7, on a pour toute $\mu \in \mathcal{M}_b$ et toute $f \in \mathcal{L}^1_+(\lambda)$ (ou toute $f \in \mathcal{L}^1(\lambda)$ par différence),

$$\int g_\sigma * f(x)\, d\mu(x) = (2\pi)^{-d/2} \int \mathrm{e}^{-i<y,u>} g_1(\sigma u)\hat{\mu}(u) f(y)\, dudy. \tag{1.15}$$

Nous pouvons maintenant établir l'injectivité de la transformation de Fourier.

Théorème 1.6.8. *(i) Soient $\mu, \nu \in \mathcal{M}_b$. Si, pour tout t, $\hat{\mu}(t) = \hat{\nu}(t)$, on a $\mu = \nu$.*
(ii) Soit $\mu \in \mathcal{M}_b$. Si $\hat{\mu} \in \mathcal{L}^1(\lambda)$, on a $\mu = h.\lambda$ avec

$$h(y) = (2\pi)^{-d} \int e^{-i<y,u>} \hat{\mu}(u)\, du.$$

`Preuve:`
(i) Vu (1.15), on a, pour toute $f \in C_k$, $\int g_\sigma * f(x)\, d\mu(x) = \int g_\sigma * f(x)\, d\nu(x)$ et donc (lem.1.6.5) $\int f(x)\, d\mu(x) = \int f(x)\, d\nu(x)$ et (lem.1.5.3) $\mu = \nu$.

(ii) Considérons (1.15) pour $f \in C_k$. Lorsque $\sigma \to 0$, le terme de gauche tend vers $\int f(x)\, d\mu(x)$ (lem.1.6.5) et $\mathrm{e}^{-i<y,u>} g_1(\sigma u)\hat{\mu}(u) f(y) \to (2\pi)^{-d/2} \mathrm{e}^{-i<y,u>} \hat{\mu}(u) f(y)$ en restant borné par $|\hat{\mu}(u) f(y)| \in \mathcal{L}^1(\lambda \otimes \lambda)$. On peut donc appliquer le th. de Lebesgue au terme de droite et on obtient à la limite,

$$\int f(x)\, d\mu(x) = (2\pi)^{-d} \int \mathrm{e}^{-i<y,u>} \hat{\mu}(u) f(y)\, dudy$$

$$= \int ((2\pi)^{-d} \int \mathrm{e}^{-i<y,u>} \hat{\mu}(u)\, du)\, f(y)\, dy = \int f(y) h(y)\, dy$$

donc (lem.1.5.3) $\mu = h.\lambda$ ∎

1.7. Convergences de mesures

Soient $\mu_n, \mu \in \mathcal{M}_b$. On veut donner un sens à "μ_n converge vers μ". La première idée est de dire que μ_n converge vers μ si, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $\mu_n(A) \to \mu(A)$. Mais ceci est très restrictif. Par exemple si $\mu_n = \delta_{\frac{1}{n}} \in \mathcal{M}_b(\mathbb{R})$, on n'a pas la convergence en ce sens de μ_n vers $\mu = \delta_0$ puisque $\mu_n(]0,1]) = 1$ et que $\mu(]0,1]) = 0$. On va plutôt définir la convergence en utilisant la dualité avec certains espaces de fonctions continues.

Définition 1.7.1. *Soient $\mu_n, \mu \in \mathcal{M}_b$. On dit que μ_n converge vers μ vaguement si, pour toute $f \in C_k$, $\int f\,d\mu_n \to \int f\,d\mu$, faiblement si, pour toute $f \in C_0$, $\int f\,d\mu_n \to \int f\,d\mu$, étroitement si, pour toute $f \in C_b$, $\int f\,d\mu_n \to \int f\,d\mu$.*

Evidemment la convergence étroite implique la convergence faible et la convergence faible implique la convergence vague. Par contre, sur $\mathbb{R}$, δ_n converge faiblement vers 0 et pas étroitement, $n\delta_n$ converge vaguement vers 0 et pas faiblement. En fait pour passer de la convergence vague à la convergence étroite, il faut ajouter la "conservation de la masse". C'est ce que montre la proposition suivante. Rappelons que, pour $\mu \in \mathcal{M}_b$ et $f \in C_b$, on écrit aussi bien $\int f\,d\mu$ que $\mu(f)$.

Proposition 1.7.2. *Soient $\mu_n, \mu \in \mathcal{M}_b$. Si μ_n converge vaguement vers μ et si $\mu_n(1) \to \mu(1)$, μ_n converge étroitement vers μ.*

`Preuve:`
On a, pour $f \in C_b$ et $g \in C_k$, $0 \le g \le 1$,

$$\begin{aligned} |\mu_n(f) - \mu(f)| &\le |\mu_n(f) - \mu_n(fg)| + |\mu_n(fg) - \mu(fg)| + |\mu(fg) - \mu(f)| \\ &\le ||f||(\mu_n(1) - \mu_n(g)) + |\mu_n(fg) - \mu(fg)| + ||f||(\mu(1) - \mu(g)). \end{aligned}$$

On a donc $\overline{\lim}_n |\mu_n(f) - \mu(f)| \le 2||f||(\mu(1) - \mu(g))$. Vu qu'il existe $g_n \in C_k$, $0 \le g_n \le 1$, tels que $g_n \uparrow 1$ et qu'alors $\mu(g_n) \uparrow \mu(1)$, $\mu(1) - \mu(g)$ est arbitrairement petit et $\mu_n(f) \to \mu(f)$. Ceci montre que μ_n converge étroitement vers μ ∎

Convergence faible

On dit que $H \subset C_0$ est total dans C_0 si l'ensemble des combinaisons linéaires d'éléments de H est dense dans C_0.

Proposition 1.7.3. *Soient $\mu_n, \mu \in \mathcal{M}_b$ et H un ensemble total dans C_0. On suppose que, pout tout $g \in H$, $\mu_n(g) \to \mu(g)$ et que $A = \sup_n \mu_n(1) < +\infty$, alors μ_n converge faiblement vers μ.*

`Preuve:`
Soit V l'espace vectoriel engendré par H. On a $\overline{V} = C_0$ et, pour toute $g \in V$, $\mu_n(g) \to \mu(g)$. Soient $f \in C_0$ et $g \in V$, on a

$$\begin{aligned} |\mu_n(f) - \mu(f)| &\le |\mu_n(f) - \mu_n(g)| + |\mu_n(g) - \mu(g)| + |\mu(g) - \mu(f)| \\ &\le ||f - g||\mu_n(1) + |\mu_n(g) - \mu(g)| + ||f - g||\mu(1). \end{aligned}$$

On a donc $\overline{\lim}_n|\mu_n(f)-\mu(f)| \leq ||f-g||(A+\mu(1))$. Cette dernière quantité étant arbitrairement petite, $\mu_n(f)\to\mu(f)$ et μ_n converge faiblement vers μ ∎

Le théorème suivant montre que l'ensemble $\{\mu\in\mathcal{M}_b,\ \mu(1)\leq M\}$ est faiblement compact.

Théorème 1.7.4. *Soient $\mu_n\in\mathcal{M}_b$ telles que $A=\sup_n\mu_n(1)<+\infty$. Il existe une sous-suite μ_{n_k} et $\mu\in\mathcal{M}_b$ telles que μ_{n_k} converge faiblement vers μ.*

`Preuve:`
Soit $V=\{\phi_1,\phi_2,\ldots,\phi_k,\ldots\}$ un ensemble dénombrable dense dans C_0. On a, pour tous n et k, $|\mu_n(\phi_k)|\leq A||\phi_k||$. La suite $\mu_n(\phi_1)$ étant bornée, il existe (n_k^1) telle que $\mu_{n_k^1}(\phi_1)$ converge; la suite $\mu_{n_k^1}(\phi_2)$ étant bornée, il existe $(n_k^2)\subset(n_k^1)$ telle que $\mu_{n_k^2}(\phi_2)$ converge;... La suite diagonale (n_k^k) est telle que $\mu_{n_k^k}(\phi)$ converge pour toute $\phi\in V$. On pose $\mu'_k=\mu_{n_k^k}$. Pour $f\in C_0$ et $\phi\in V$, on a

$$\begin{aligned}|\mu'_{n+k}(f)-\mu'_n(f)| &\leq |\mu'_{n+k}(f)-\mu'_{n+k}(\phi)|+|\mu'_{n+k}(\phi)-\mu'_n(\phi)|+|\mu'_n(\phi)-\mu'_n(f)|\\ &\leq ||f-\phi||\mu'_{n+k}(1)+|\mu'_{n+k}(\phi)-\mu'_n(\phi)|+||f-\phi||\mu'_n(1).\end{aligned}$$

D'où $\overline{\lim}_n\sup_k|\mu'_{n+k}(f)-\mu'_n(f)|\leq 2A||f-\phi||$. Cette dernière quantité étant arbitrairement petite, $\mu'_n(f)$ est une suite de Cauchy et donc converge pour toute $f\in C_0$. On pose $I(f)=\lim_n\mu'_n(f)$. Alors $f\mapsto I(f)$ est une forme linéaire positive sur C_0, il existe donc (th.1.5.4) $\mu\in\mathcal{M}_b$ telle que $I(f)=\mu(f)$ et μ'_k converge faiblement vers μ ∎

Convergence étroite

Pour le probabiliste, c'est la plus intéressante puisqu'elle conserve les probabilités! Revenons d'abord au problème de la convergence sur les ensembles. On a vu que μ_n peut converger étroitement vers μ sans que $\mu_n(A)$ converge vers $\mu(A)$. La question est de savoir pour quels ensembles on a cette convergence. On note $\partial A=\overline{A}\setminus\mathring{A}$ la frontière topologique de A i.e. la fermeture moins l'intérieur.

Proposition 1.7.5. *Soient $\mu_n,\mu\in\mathcal{M}_b$. On suppose que μ_n converge étroitement vers μ. Alors, pour tout $A\in\mathcal{B}(\mathbb{R}^d)$ tel que $\mu(\partial A)=0$, $\mu_n(A)\to\mu(A)$.*

`Preuve:`
Il existe $f_p,g_p\in C_b^+$ telles que $g_p\downarrow \mathbb{1}_{\{\overline{A}\}}$, $f_p\uparrow\mathbb{1}_{\{\mathring{A}\}}$, alors $\mu(g_p)\downarrow\mu(\overline{A})$ et $\mu(f_p)\uparrow\mu(\mathring{A})$. On a donc, vu l'hypothèse, $\mu(g_p-f_p)\to_p 0$ et, vu que $f_p\leq\mathbb{1}_{\{A\}}\leq g_p$,

$$\mu_n(f_p)-\mu(g_p)\leq\mu_n(A)-\mu(A)\leq\mu_n(g_p)-\mu(f_p)$$

d'où $|\mu_n(A)-\mu(A)|\leq|\mu_n(g_p)-\mu(f_p)|+|\mu_n(f_p)-\mu(g_p)|$. On a donc $\overline{\lim}_n|\mu_n(A)-\mu(A)|\leq 2(\mu(g_p)-\mu(f_p))$. Cette dernière quantité étant arbitrairement petite, $\mu_n(A)\to\mu(A)$ ∎

Le th.1.7.4 n'est pas valable pour la convergence étroite car la "masse" de μ_n peut aller à l'infini (il suffit de considérer $\mu_n=\delta_n$). On va préciser ce point.

Lemme 1.7.6. *Soit $\mu_n \in \mathcal{M}_b$ une suite convergeant étroitement vers μ, alors, pour tout $\varepsilon > 0$, il existe un compact K tel que, pour tout n, $\mu_n(K^c) < \varepsilon$.*

`Preuve:`
Pour a assez grand, $\mu(|x| > a) < \varepsilon$. Soit $f \in C_b$ telle que $\mathbb{1}_{\{\{|x|>a+1\}\}} \leq f \leq \mathbb{1}_{\{\{|x|>a\}\}}$. On a $\mu(f) < \varepsilon$ et donc, pour tout $n \geq n_0$, $\mu_n(f) < \varepsilon$ d'où $\mu_n(\{|x| \leq a+1\}^c) < \varepsilon$. On conclut facilement ∎

En fait cette condition suffit à assurer la compacité étroite, c'est le théorème de Prokhorov.

Définition 1.7.7. *Une suite $\mu_n \in \mathcal{M}_b$ est dite tendue si, pour tout $\varepsilon > 0$, il existe un compact K tel que, pour tout n, $\mu_n(K^c) < \varepsilon$.*

Théorème 1.7.8. *Soit $\mu_n \in \mathcal{M}_b$ une suite tendue vérifiant $A = \sup_n \mu_n(1) < +\infty$. Il existe une sous-suite μ_{n_k} convergeant étroitement.*

`Preuve:`
On sait (th.1.7.4) qu'il existe une sous-suite $\mu'_k = \mu_{n_k}$ convergeant faiblement vers $\mu \in \mathcal{M}_b$. Soit $\varepsilon > 0$. Il existe, c'est l'hypothèse, un compact K tel que, pour tout n, $\mu'_n(K^c) < \varepsilon$ et aussi, quitte à agrandir K, $\mu(K^c) < \varepsilon$. Soit $f \in C_k$, $0 \leq f \leq 1$, $f \geq \mathbb{1}_{\{K\}}$. On a $0 \leq 1 - f \leq \mathbb{1}_{\{K^c\}}$ et

$$|\mu'_n(1) - \mu(1)| \leq |\mu'_n(f) - \mu(f)| + \mu'_n(1-f) + \mu(1-f) \leq |\mu'_n(f) - \mu(f)| + 2\varepsilon.$$

Donc $\overline{\lim}_n |\mu'_n(1) - \mu(1)| \leq 2\varepsilon$ et $\mu'_n(1) \to \mu(1)$. Vu la prop.1.7.2, μ'_n converge étroitement vers μ ∎

Le critère suivant est essentiel en probabilités.

Théorème 1.7.9. *Soient $\mu_n, \mu \in \mathcal{M}_b$. Alors μ_n converge étroitement vers μ ssi, pour tout $t \in \mathbb{R}^d$, $\hat{\mu}_n(t) \to \hat{\mu}(t)$.*

`Preuve:`
Puisque $t \mapsto \mathrm{e}^{i<t,x>}$ est continue bornée, μ_n converge étroitement vers μ implique que $\hat{\mu}_n(t) \to \hat{\mu}(t)$. Réciproquement supposons que, pour tout $t \in \mathbb{R}^d$, $\hat{\mu}_n(t) \to \hat{\mu}(t)$. On a en particulier $\mu_n(1) = \hat{\mu}_n(0) \to \hat{\mu}(0) = \mu(1)$ d'où $A = \sup_n \mu_n(1) < +\infty$ et $|\hat{\mu}_n(t)| \leq |\hat{\mu}_n(0)| \leq A$. Soit $f \in C_k$, on a, vu (1.15),

$$\int g_\sigma * f(x)\, d\mu_n(x) = (2\pi)^{-d/2} \int \mathrm{e}^{-i<y,u>} g_1(\sigma u) \hat{\mu}_n(u) f(y)\, dudy.$$

Par hypothèses $\mathrm{e}^{-i<y,u>} g_1(\sigma u) \hat{\mu}_n(u) f(y) \to \mathrm{e}^{-i<y,u>} g_1(\sigma u) \hat{\mu}(u) f(y)$, son module étant borné par $A g_1(\sigma u)|f(y)| \in \mathcal{L}^1(dudy)$. On a donc (th. de Lebesgue)

$$(2\pi)^{-d/2} \int \mathrm{e}^{-i<y,u>} g_1(\sigma u) \hat{\mu}_n(u) f(y)\, dudy \to_n (2\pi)^{-d/2} \int \mathrm{e}^{-i<y,u>} g_1(\sigma u) \hat{\mu}(u) f(y)\, dudy.$$

Ce dernier terme est égal, vu (1.15), à $\int g_\sigma * f(x)\, d\mu(x)$. Donc, pour toute $f \in C_k$, $\int g_\sigma * f(x)\, d\mu_n(x) \to \int g_\sigma * f(x)\, d\mu(x)$. Les $(g_\sigma,\ \sigma > 0,\ f \in C_k)$ étant denses (cor.1.6.6)

dans C_0, μ_n converge faiblement vers μ (prop.1.7.3). Comme $\mu_n(1) \to \mu(1)$, μ_n converge étroitement vers μ (prop.1.7.2) ∎

S'il est souvent facile de calculer la limite $\phi(t)$ de $\hat{\mu}_n(t)$, il est plus délicat de voir si ϕ est une transformée de Fourier. Le théorème de P. Lévy répond à cette question.

Théorème 1.7.10. *Soient $\mu_n \in \mathcal{M}_b$ telles que, pour tout $t \in \mathbb{R}^d$, $\hat{\mu}_n(t) \to \phi(t)$. Si la fonction ϕ est continue en 0, il existe $\mu \in \mathcal{M}_b$ telle que $\phi = \hat{\mu}$ et μ_n converge étroitement vers μ.*

`Preuve:`
Vu que $\mu_n(1) = \hat{\mu}_n(0) \to \phi(0)$, on a $\sup_n \mu_n(1) = A < +\infty$. De plus $|\hat{\mu}_n(t)| \leq |\hat{\mu}_n(0)| \leq A$ et $|\phi(t)| \leq A$. Il existe donc (th.1.7.4) une sous-suite μ_{n_k} telle que μ_{n_k} converge faiblement vers $\mu \in \mathcal{M}_b$. On pose $\mu'_k = \mu_{n_k}$. D'après (1.14), on a, pour tout $y \in \mathbb{R}^d$,

$$\int g_\sigma(x-y)\, d\mu'_k(x) = (2\pi)^{-d/2} \int \mathrm{e}^{-i<y,u>} g_1(\sigma u)\hat{\mu}'_k(u)\, du.$$

Passant à la limite en k, on a (la justification est la même qu'au cours de la preuve du th.1.7.9),

$$\int g_\sigma(x-y)\, d\mu(x) = (2\pi)^{-d/2} \int \mathrm{e}^{-i<y,u>} g_1(\sigma u)\phi(u)\, du.$$

On a donc vu (1.14)

$$\int \mathrm{e}^{-i<y,u>} g_1(\sigma u)\hat{\mu}(u)\, du = \int \mathrm{e}^{-i<y,u>} g_1(\sigma u)\phi(u)\, du.$$

D'où (th.1.6.8) $\hat{\mu}(u)g_1(\sigma u) = \phi(u)g_1(\sigma u)$ λ p.p. et, g_1 étant > 0, $\hat{\mu}(u) = \phi(u)$ λ p.p. Soit $E = \{\hat{\mu} = \phi\}$, on a $\lambda(E^c) = 0$. Il existe donc $x_n \in E$ tel que $x_n \to 0$. On a, pour tout n, $\hat{\mu}(x_n) = \phi(x_n)$ et, les deux fonctions étant continues en 0, $\hat{\mu}(0) = \phi(0)$. On a donc $\mu'_k(1) = \hat{\mu}'_k(0) \to \phi(0) = \hat{\mu}(0) = \mu(1)$ et (prop.1.7.2) μ'_k converge étroitement vers μ. On en déduit (th.1.7.9) que $\phi = \hat{\mu}$ et que μ_n converge étroitement vers μ ∎

1.8. Mesures signées

Soit $(E, \mathcal{B})$ un espace mesurable.

On a réservé le nom de "mesures" aux applications σ-additives de $\mathcal{B}$ dans $\overline{\mathbb{R}}^+$ mais on peut aussi s'intéresser aux applications σ-additives de $\mathcal{B}$ dans $\mathbb{R}$.

Définition 1.8.1. *On appelle mesure signée toute application μ de $\mathcal{B}$ dans $\mathbb{R}$ telle que*
(i) $\mu(\emptyset) = 0$,
(ii) pour tous $A_n \in \mathcal{B}$ deux à deux disjoints et de réunion A, la série $\sum \mu(A_n)$ converge et $\sum_n \mu(A_n) = \mu(A)$.

Remarquons qu'une mesure bornée est une mesure signée mais qu'une mesure non bornée n'est pas une mesure signée à cause de la valeur $+\infty$.

Exemple. Soit $\mu = \mu_1 - \mu_2$ avec μ_1, μ_2 mesures bornées, alors μ est une mesure signée (on verra que toutes les mesures signées sont obtenues ainsi). Si μ_1 et μ_2 ont des densités h_1 et h_2 par rapport à une mesure λ, on a $\mu(A) = \int_A h\,d\lambda$ avec $h = h_1 - h_2 \in \mathcal{L}^1(\lambda)$. Si on pose $S = \{h \geq 0\}$, on a, pour tout A, $\mu(A \cap S) = \int_A h \mathbb{1}_{\{S\}}\,d\lambda \geq 0$ et $\mu(A \cap S^c) = \int_A h \mathbb{1}_{\{S^c\}}\,d\lambda \leq 0$ i.e. S "porte" la partie positive de μ et S^c "porte" la partie négative de μ. En fait cette situation est représentative du cas général et on peut montrer:

Théorème 1.8.2. *Soit μ une mesure signée sur $(E, \mathcal{B})$. Il existe $S \in \mathcal{B}$ tel que, pour tout $A \in \mathcal{B}$, $\mu(A \cap S) \geq 0$, $\mu(A \cap S^c) \leq 0$.*

On pose

$$\mu_+(A) = \mu(A \cap S), \quad \mu_-(A) = -\mu(A \cap S^c), \quad |\mu| = \mu_+ + \mu_-. \tag{1.16}$$

Alors $\mu_+, \mu_-, |\mu|$ sont des mesures bornées sur $(E, \mathcal{B})$ appelées partie positive, partie négative et valeur absolue de μ. On a évidemment $\mu = \mu_+ - \mu_-$. Donnons en quelques propriétés simples.

Corollaire 1.8.3. *$|\mu|$ est la plus petite mesure telle que, pour tout $A \in \mathcal{B}$, $\nu(A) \geq |\mu(A)|$.*

`Preuve:`
D'une part $|\mu(A)| = |\mu(A \cap S) + \mu(A \cap S^c)| \leq |\mu(A \cap S)| + |\mu(A \cap S^c)| = |\mu|(A \cap S) + |\mu|(A \cap S^c) = |\mu|(A)$. D'autre part si, pour tout A, $\nu(A) \geq |\mu(A)|$, on a $\nu(A) = \nu(A \cap S) + \nu(A \cap S^c) \geq |\mu(A \cap S)| + |\mu(A \cap S^c)| = |\mu|(A \cap S) + |\mu|(A \cap S^c) = |\mu|(A)$ ∎

Corollaire 1.8.4. *Si $\mu = \nu_1 - \nu_2$, ν_1, ν_2 mesures bornées, on a $\nu_1 \geq \mu_+$ et $\nu_2 \geq \mu_-$.*

`Preuve:`
On a $\mu_+(A) = \mu(A \cap S) \leq \nu_1(A \cap S) \leq \nu_1(A)$ et $\mu_-(A) = -\mu(A \cap S^c) \leq \nu_2(A \cap S^c) \leq \nu_2(A)$ ∎

Corollaire 1.8.5. *On a, pour tout $A \in \mathcal{B}$, $\mu(A) = \int_A h\,d|\mu|$ avec $|h| = 1$.*

`Preuve:`
On a $\mu(A) = \mu(A \cap S) + \mu(A \cap S^c) = |\mu|(A \cap S) - |\mu|(A \cap S^c) = \int_A (\mathbb{1}_{\{S\}} - \mathbb{1}_{\{S^c\}})\,d|\mu|$ ∎

Théorème de Radon-Nikodym

Si $\mu(A) = \int_A h\,d\lambda$ est une mesure signée, on a évidemment, pour $A \in \mathcal{B}$, $\lambda(A) = 0$ implique $\mu(A) = 0$. Il est remarquable que cette propriété suffise à caractériser les mesures ayant une densité par rapport à λ.

Définition 1.8.6. *Soit λ une mesure sur $(E, \mathcal{B})$. On dit qu'une mesure signée sur $(E, \mathcal{B})$ (ou σ-finie) μ est absolument continue par rapport à λ si*

$$A \in \mathcal{B} \text{ et } \lambda(A) = 0 \text{ impliquent } \mu(A) = 0.$$

On note alors $\mu \ll \lambda$. On a:

Théorème 1.8.7. *On considère un espace mesuré $(E, \mathcal{B}, \lambda)$, λ σ-finie.*
(i) Soit μ une mesure signée sur $(E, \mathcal{B})$ telle que $\mu \ll \lambda$. Il existe $\phi \in \mathcal{L}^1(\lambda)$, unique à un λ p.p. près, telle que, pour tout $A \in \mathcal{B}$, $\mu(A) = \int_A \phi \, d\lambda$.
(ii) Soit μ une mesure σ-finie sur $(E, \mathcal{B})$ telle que $\mu \ll \lambda$. Il existe $\phi \in \mathcal{B}^+$, unique à un λ p.p. près, telle que, pour tout $A \in \mathcal{B}$, $\mu(A) = \int_A \phi \, d\lambda$.

`Preuve:`
(i) Supposons d'abord μ, λ mesures bornées et $\mu \ll \lambda$. On pose $\rho = \lambda + \mu$. On a $\mathcal{L}^2(\rho) \subset \mathcal{L}^1(\rho) \subset \mathcal{L}^1(\mu)$ et $f = g$ ρ p.p. implique $f = g$ μ p.p. On peut donc définir $I(f) = \int f \, d\mu$ pour $f \in L^2(\rho)$ et l'on a

$$|I(f)| \leq \int |f| \, d\mu \leq \int |f| \, d\rho \leq (\rho(E))^{1/2} (\int |f|^2 \, d\rho)^{1/2}.$$

Ceci montre que $f \mapsto I(f)$ est une forme linéaire continue sur $L^2(\rho)$. Il existe donc $g \in L^2(\rho)$ telle que $I(f) =< g, f >$ et donc $g \in \mathcal{L}^2(\rho)$ telle

$$\text{pour toute } f \in \mathcal{L}^2(\rho), \int f \, d\mu = \int f g \, d\rho. \tag{1.17}$$

Prenant $f = \mathbb{1}_{\{A\}}$ dans (1.17), on a, pour tout $A \in \mathcal{B}$,

$$0 \leq \int_A g \, d\rho = \mu(A) \leq \rho(A) = \int_A 1 \, d\rho,$$

ce qui implique que $0 \leq g \leq 1$ ρ p.p. Prenant $f = \mathbb{1}_{\{\{g=1\}\}}$ dans (1.17), on a $\mu(g = 1) = \rho(g = 1)$ d'où $\lambda(g = 1) = 0$ et, vu l'hypothèse, $\mu(g = 1) = 0$ et enfin $\rho(g = 1) = 0$. On a donc $0 \leq g < 1$ ρ p.p.

De (1.17) on tire, puisque $\rho = \lambda + \mu$, pour toute $f \in \mathcal{L}^2(\rho)$, $\int f(1 - g) \, d\mu = \int f g \, d\lambda$. Par limite croissante, on a

$$\text{pour toute } f \in \mathcal{B}^+, \int f(1 - g) \, d\mu = \int f g \, d\lambda. \tag{1.18}$$

Prenant $f = \dfrac{\mathbb{1}_{\{A\}}}{1 - g}$ dans (1.18), on a $\mu(A) = \int_A \phi \, d\lambda$ avec $\phi = \dfrac{g}{1 - g}$.

(ii) Supposons maintenant μ mesure signée, λ mesure bornée et $\mu \ll \lambda$. Puisque $|\mu|(A) = \mu(A \cap S) - \mu(A \cap S^s) = 0$ si $\lambda(A) = 0$, on a $|\mu| \ll \lambda$. Donc $|\mu| = \phi.\lambda$ et (cor.1.8.5) $\mu = \phi h.\lambda$.

(iii) Ces résultats s'étendent immédiatement au cas σ-fini en "découpant" E.

Enfin l'unicité résulte facilement de la propriété (iv) 1.3.3. ∎

Soit μ une mesure signée sur $(E, \mathcal{B})$. On a (cor.1.8.5), $\mu = h.|\mu|$ avec $|h| = 1$. Si $f \in \mathcal{L}(|\mu|)$, on définit

$$\int f \, d\mu := \int f h \, d|\mu|.$$

Revenons au cas où $E = \mathbb{R}^d$. On note $\mathcal{M}_s$ l'ensemble des mesures signées sur $\mathcal{B}(\mathbb{R}^d)$. Si $\mu \in \mathcal{M}_s$ et $f \in C_0 = C_0(\mathbb{R}^d)$, posons $I(f) = \int f\,d\mu$. On a $|I(f)| = |\int fh\,d|\mu|| \leq \int |f|\,d|\mu| \leq |\mu|(\mathbb{R}^d)||f||$. Donc $f \mapsto I(f)$ est une forme linéaire continue sur C_0. En fait, elles sont toutes de cette forme:

Théorème 1.8.8. *Soit $f \mapsto I(f)$ une forme linéaire continue sur C_0. Il existe $\mu \in \mathcal{M}_s$ unique telle que, pour toute $f \in C_0$, $I(f) = \int f\,d\mu$.*

Chapitre 2

Notions de probabilités

2.1. Espace de probabilité

Tout commence par:

Définition 2.1.1. *On appelle espace de probabilité un triplet* $(\Omega, \mathcal{A}, \mathbb{P})$ *où* $(\Omega, \mathcal{A})$ *est un espace mesurable et* $\mathbb{P}$ *une probabilité sur* $\mathcal{A}$*.*

Les éléments de $\mathcal{A}$ s'appellent des événements. Pour des événements A et B, on écrira indifféremment $A \cap B$ ou AB.

Premières propriétés. A_n, A, B étant des événements,

(i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$,

(ii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, $\mathbb{P}(\cup A_n) \leq \sum \mathbb{P}(A_n)$

(iii) si $A_n \uparrow A$, $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$, si $A_n \downarrow A$, $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$

Rappelons qu'un sous-ensemble B de Ω est dit négligeable si $B \subset A \in \mathcal{A}$ tel que $\mathbb{P}(A) = 0$. Une propriété dépendant de ω est vraie presque sûrement, en abrégé p.s., si elle est vraie en dehors d'un ensemble négligeable. Notons qu'un ensemble négligeable n'est pas toujours un événement sauf si l'espace $(\Omega, \mathcal{A}, \mathbb{P})$ est complet. On peut cependant toujours se ramener à ce cas. Voir à ce sujet 1.2.3.

Probabilité conditionnelle.

Définition 2.1.2. *Soient* $A, B \in \mathcal{A}$ *avec* $P(B) > 0$*. On appelle probabilité conditionnelle de* A *sachant* B *et on note* $\mathbb{P}(A|B)$ *la quantité* $\mathbb{P}(A \cap B)/\mathbb{P}(B)$*.*

Noter que $A \mapsto \mathbb{P}(A|B)$ est une probabilité sur $(\Omega, \mathcal{A})$. La proposition suivante s'appelle la formule de Bayes.

Proposition 2.1.3. *Soient* $(B_n,\, n \in \mathbb{N})$ *une partition de* Ω *avec* $B_n \in \mathcal{A}$ *et* $\mathbb{P}(B_n) > 0$*. On a, pour tout* $A \in \mathcal{A}$ *tel que* $\mathbb{P}(A) > 0$ *et tout* n*,*

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(B_n)\mathbb{P}(A|B_n)}{\sum_k \mathbb{P}(B_k)\mathbb{P}(A|B_k)}$$

`Preuve:`
On a $\mathbb{P}(B_n|A) = \frac{\mathbb{P}(A \cap B_n)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_n)\mathbb{P}(A|B_n)}{\mathbb{P}(A)}$ et $\mathbb{P}(A) = \sum_k \mathbb{P}(A \cap B_k) = \sum_k \mathbb{P}(B_k)\mathbb{P}(A|B_k)$
∎

Proposition 2.1.4. *Soient $A_1, \ldots, A_n$ des événements tels que $\mathbb{P}(A_1 \ldots A_n) > 0$. Alors*

$$\mathbb{P}(A_1 \ldots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1A_2) \ldots \mathbb{P}(A_n|A_1 \ldots A_{n-1}).$$

`Preuve:`
Par récurrence ∎

Si $\mathbb{P}(A|B) = \mathbb{P}(A)$, l'occurrence de A n'est pas influencée par celle de B, on dit que les événements A et B sont indépendants. Ceci s'écrit aussi $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On voit facilement qu'alors A et B^c, A^c et B, A^c et B^c sont aussi indépendants. En fait ce sont les tribus $\sigma(A) = \{\Omega, \emptyset, A, A^c\}$ et $\sigma(B) = \{\Omega, \emptyset, B, B^c\}$ qui sont indépendantes. Nous développerons cette notion en 2.2.

Variables aléatoires.

Les variables aléatoires sont les fonctions qui dépendent du hasard, celui-ci étant modélisé par le tirage d'un point $\omega \in \Omega$.

Définition 2.1.5. *On appelle variable aléatoire (en abrégé v.a.) à valeurs $(E, \mathcal{E})$ toute application mesurable de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{E})$.*

Si E est dénombrable et $\mathcal{E} = \mathcal{P}(E)$, on parle de v.a. discrète,

si $E = \overline{\mathbb{R}}^+$ et $\mathcal{E} = \mathcal{B}(\overline{\mathbb{R}}^+)$, on parle de v.a. positive,

si $E = \mathbb{R}$ et $\mathcal{E} = \mathcal{B}(\mathbb{R})$, on parle de v.a. réelle (v.a.r.),

si $E = \mathbb{R}^d$ et $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$, on parle de v.a. vectorielle,

si $E = \mathbb{C}$ et $\mathcal{E} = \mathcal{B}(\mathbb{C})$, on parle de v.a. complexe.

Définition 2.1.6. *Soit X une v.a.*

- *Si X est positive, on appelle espérance de X et on note $\mathbb{E}(X)$ la quantité $\int X \, d\mathbb{P}$.*
- *Si X est réelle, telle que $\mathbb{E}|X| < +\infty$, on appelle espérance de X et on note $\mathbb{E}(X)$ la quantité $\int X \, d\mathbb{P}$.*

Définition 2.1.7. *Soit X une v.a. à valeurs $(E, \mathcal{E})$. On appelle loi de X et on note μ_X la mesure image (prop.1.3.11) de $\mathbb{P}$ par X.*

Il résulte de la prop.1.3.11 que μ_X est la probabilité sur $(E, \mathcal{E})$ définie par

$$\mu_X(\Gamma) = \mathbb{P}(X \in \Gamma) \text{ où } \{X \in \Gamma\} = \{\omega \in \Omega, \ X(\omega) \in \Gamma\} = X^{-1}(\Gamma), \tag{2.1}$$

et que

$$\text{pour toute } f \in \mathcal{E}^+ \cup \mathcal{L}^1(E, \mathcal{E}, \mu_X), \ \mathbb{E}(f(X)) = \int f \, d\mu_X \,. \tag{2.2}$$

Proposition 2.1.8. *Soient X une v.a. à valeurs $(E,\mathcal{E})$ et ϕ une application mesurable de $(E,\mathcal{E})$ dans $(F,\mathcal{F})$, alors $Y=\phi(X)$ est une v.a. à valeurs $(F,\mathcal{F})$ et la loi de Y est l'image par ϕ de la loi de X.*

`Preuve:`
Le premier point résulte de ce que la composée de deux applications mesurables est mesurable. Quant au second, on a, pour $f \in \mathcal{F}^+$,

$$\int f\,d\mu_Y = \mathbb{E}(f(Y)) = \mathbb{E}(f(\phi(X))) = \int f \circ \phi\,d\mu_X$$

i.e. (prop.1.3.11) μ_Y est l'image par ϕ de μ_X ∎

Exemples. Il y a deux situations fondamentales.

(i) X est discrète i.e. E est dénombrable. La loi μ_X est alors déterminée par la famille $(\mu_X(x),\ x \in E)$ où $\mu_X(x) = \mathbb{P}(X=x)$ et l'on a

$$\text{pour toute } f \geq 0, \quad \mathbb{E}(f(X)) = \sum_{x \in E} f(x)\mu_X(x).$$

(ii) X est vectorielle i.e. à valeurs $\mathbb{R}^d$ et $\mu_X = h_X.\lambda$, λ étant la mesure de Lebesgue sur $\mathbb{R}^d$. On dit alors que X est une v.a. de densité h_X. Dans ce cas, on a,

$$\text{pour toute } f \in \mathcal{B}^+(\mathbb{R}^d), \quad \mathbb{E}(f(X)) = \int f h_X\,d\lambda.$$

2.2. Indépendance

Toutes les tribus considérées sont des sous-tribus de $\mathcal{A}$.

Tribus indépendantes.

Définition 2.2.1. *Des tribus $\mathcal{B}_i$, $i=1,\ldots,n$ sont dites indépendantes si*

$$\textit{pour tous } A_i \in \mathcal{B}_i\,, \quad \mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i).$$

Une famille quelconque $(\mathcal{B}_i,\ i \in I)$ de tribus est dite indépendante si toute sous famille finie est indépendante.

Cette définition a comme conséquence évidente mais importante:

Lemme 2.2.2. *Si les tribus $(\mathcal{B}_i,\ i \in I)$ sont indépendantes et si, pour chaque $i \in I$, $\mathcal{B}'_i$ est une sous-tribu de $\mathcal{B}_i$, les tribus $(\mathcal{B}'_i,\ i \in I)$ sont indépendantes.*

Proposition 2.2.3. *Soient $\mathcal{C}_k \subset \mathcal{P}(\Omega)$, $k = 1, \ldots, n$ des classes stables par intersection finie et contenant Ω. On suppose que*

$$\text{pour tous } A_k \in \mathcal{C}_k,\ k = 1, \ldots, n,\ \mathbb{P}(\cap_{k=1}^n A_k) = \prod_{k=1}^n \mathbb{P}(A_k). \tag{2.3}$$

Alors les tribus $\sigma(\mathcal{C}_k)$, $k = 1, \ldots, n$, sont indépendantes.

`Preuve:`
On considère la propriété pour $k = 1, \ldots, n$,

$$(P_k) : \forall A_i \in \sigma(\mathcal{C}_i),\ i = 1, \ldots, k-1,\ \forall A_i \in \mathcal{C}_i,\ i = k, \ldots, n,\ \mathbb{P}(\cap_{k=1}^n A_k)) = \prod_{k=1}^n \mathbb{P}(A_k).$$

Par hypothèse (P_1) est vraie. Supposons (P_r). On pose

$$\mathcal{M} = \{B,\ \forall A_i \in \sigma(\mathcal{C}_i),\ i = 1, \ldots, r-1,\ \forall A_i \in \mathcal{C}_i,\ i = r+1, \ldots, n,$$
$$\mathbb{P}(A_1 \ldots A_{r-1} B A_{r+1} \ldots A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_{r-1}) \mathbb{P}(B) \mathbb{P}(A_{r+1}) \ldots \mathbb{P}(A_n)\}.$$

$\mathcal{M}$ contient Ω, est stable par limite croissante et contient, vu (P_r), $\mathcal{C}_r$ donc (prop.1.1.1) $\mathcal{M}$ contient $\sigma(\mathcal{C}_r)$ et (P_{r+1}) est vraie. On a donc (P_n) qui est le résultat cherché ∎

Enfin on a "l'indépendance par paquets":

Proposition 2.2.4. *Soient $(\mathcal{B}_i,\ i \in I)$ des tribus indépendantes et $(I_j,\ j \in J)$ une partition de I. Alors les tribus $(\sigma(\mathcal{B}_i,\ i \in I_j),\ j \in J)$ sont indépendantes.*

`Preuve:`
Il suffit de considérer le cas J fini. Soit

$$\mathcal{C}_j = \{B,\ B = A_1 A_2 \ldots A_n,\ A_k \in \cup_{i \in I_j} \mathcal{B}_i\}.$$

Vu l'indépendance des $\mathcal{B}_i$, on a, pour tous $B_j \in \mathcal{C}_j$, $\mathbb{P}(\cap B_j) = \prod \mathbb{P}(B_j)$. Mais les $\mathcal{C}_j$ sont stables par intersection finie, $\Omega \in \mathcal{C}_j$ et $\sigma(\mathcal{C}_j) = \sigma(\mathcal{B}_i,\ i \in I_j)$. On applique la prop.2.2.3 ∎

Variables aléatoires indépendantes.

Définition 2.2.5. *Des v.a. $(X_i, i \in I)$ à valeurs $(E_i, \mathcal{E}_i)$ sont dites indépendantes si les tribus $(\sigma(X_i),\ i \in I)$ sont indépendantes.*
Des événements $(A_i, i \in I)$ sont dits indépendants si les tribus $(\sigma(A_i),\ i \in I)$ sont indépendantes.

On a immédiatement,

Lemme 2.2.6. *Si les tribus $(\mathcal{B}_i,\ i \in I)$ sont indépendantes et si, pour chaque $i \in I$, X_i est une v.a. $\mathcal{B}_i$-mesurable, les v.a. $(X_i,\ i \in I)$ sont indépendantes.*

`Preuve:`
On remarque que $\sigma(X_i) \subset \mathcal{B}_i$ et on applique le lem.2.2.2 ∎

Le résultat essentiel est:

Théorème 2.2.7. *Soient X_i des v.a. à valeurs $(E_i, \mathcal{E}_i)$ $i = 1, \dots, n$. Les propriétés suivantes sont équivalentes:*

(i) *Les v.a. $X_1, \dots, X_n$ sont indépendantes.*

(ii) *Pour tous $\Gamma_i \in \mathcal{E}_i$, $\mathbb{P}(X_1 \in \Gamma_1, \dots, X_n \in \Gamma_n) = \mathbb{P}(X_1 \in \Gamma_1) \dots \mathbb{P}(X_n \in \Gamma_n)$.*

(iii) *Pour tous $\Gamma_i \in \mathcal{D}_i$, $\mathbb{P}(X_1 \in \Gamma_1, \dots, X_n \in \Gamma_n) = \mathbb{P}(X_1 \in \Gamma_1) \dots \mathbb{P}(x_n \in \Gamma_n)$ où, pour chaque i, $\mathcal{D}_i$ est une classe stable par intersection finie, contenant E_i et telle que $\sigma(\mathcal{D}_i) = \mathcal{E}_i$.*

(iv) *Pour toutes $f_i \in \mathcal{E}_i^+$ (resp. $f_i \in b\mathcal{E}_i$), $\mathbb{E}(f_1(X_1) \dots f_n(X_n)) = \mathbb{E}(f_1(X_1)) \dots \mathbb{E}(f_n(X_n))$.*

(v) $\mu_{(X_1,\dots,X_n)} = \mu_{X_1} \otimes \dots \otimes \mu_{X_n}$.

`Preuve:`
(i)⇔(ii). C'est la définition.

(ii)⇒(v). On a, pour tous $\Gamma_i \in \mathcal{E}_i$,

$$\mu_{(X_1,\dots,X_n)}(\Gamma_1 \times \dots \times \Gamma_n) = \mu_{X_1}(\Gamma_1) \dots \mu_{X_n}(\Gamma_n)$$

ce qui entraîne l'égalité des mesures (th.1.4.1).

(v)⇒(iv). C'est le th.1.4.1 ou le th.1.4.2.

(iv)⇒(iii). On prend $f_i = \mathbb{1}_{\{\Gamma_i\}}$.

(iii)⇒(ii). On applique la prop.2.2.3 aux $\mathcal{C}_i = \{X_i^{-1}(\Gamma), \Gamma \in \mathcal{D}_i\}$ ∎

Corollaire 2.2.8. *Soient $X_1, \dots, X_n$ des v.a.r., il y a équivalence entre*
(i) Les v.a. $X_1, \dots, X_n$ sont indépendantes.
(ii) Pour tous $a_i, b_i \in \mathbb{R}$,

$$\mathbb{P}(a_1 < X_1 < b_1, \dots, a_n < X_n < b_n) = \mathbb{P}(a_1 < X_1 < b_1) \dots \mathbb{P}(a_n < X_n < b_n).$$

(iii) Pour toutes f_i continues à support compact

$$\mathbb{E}(f_1(X_1) \dots f_n(X_n)) = \mathbb{E}(f_1(X_1)) \dots \mathbb{E}(f_n(X_n)).$$

`Preuve:`
Il suffit de remarquer que (iii)⇒(ii) puisque $\mathbb{1}_{\{]a,b[\}} = \lim \uparrow f_m$ avec $f_m \in C_k$ et d'appliquer le th.2.2.7 ∎

Corollaire 2.2.9. *Soient $X_1, \dots, X_n$ des v.a.r. intégrables indépendantes. Alors le produit $X_1 \dots X_n$ est intégrable et $\mathbb{E}(X_1 \dots X_n) = \mathbb{E}(X_1) \dots \mathbb{E}(X_n)$.*

`Preuve:`
On a, vu (iv), $\mathbb{E}|X_1 \ldots X_n| = \mathbb{E}|X_1| \ldots \mathbb{E}|X_n| < +\infty$. Donc $X_1 \ldots X_n \in L^1(\mu_{(X_1,\ldots,X_n)})$ et on applique (v) et le th. de Fubini ∎

Remarque 1. Si les v.a. $X_1, \ldots, X_n$ sont à valeurs E_i dénombrables, une condition nécessaire et suffisante d'indépendance est (il suffit de sommer)

$$\forall x_i \in E_i, \;\; \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n).$$

Remarque 2. Attention. On peut avoir X indépendante de Y, X indépendante de Z sans que X soit indépendante de (Y, Z). Par exemple soient X et Y deux v.a. indépendantes telles que $\mathbb{P}(X = 1) = \mathbb{P}(Y = 1) = \mathbb{P}(X = -1) = \mathbb{P}(Y = -1) = \frac{1}{2}$. On pose $Z = XY$. On a encore $\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}$. On vérifie facilement que X et Z sont indépendantes (en effet $\mathbb{P}(X = 1, Z = 1) = \mathbb{P}(X = 1, Y = 1) = \frac{1}{4} = \mathbb{P}(X = 1)\mathbb{P}(Z = 1), \ldots$). Mais X n'est pas indépendante de $Z/Y = X$ car ceci implique $X = C^{te}$. En fait la classe $\mathcal{C} = \{A, \; A = \{Y \in \Gamma_1\} \text{ ou } A = \{Z \in \Gamma_2\}\}$ n'est pas stable par intersection.

Loi 0-**1.**

Proposition 2.2.10. *Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. indépendantes. On pose*

$$\mathcal{B}_\infty = \cap_{n \geq 1} \sigma(X_k, \; k \geq n).$$

Alors, pour tout $A \in \mathcal{B}_\infty$, $\mathbb{P}(A) = 0$ ou 1. De plus, si X est une v.a.r. $\mathcal{B}_\infty$-mesurable alors X est p.s. constante.

`Preuve:`
Posons $\mathcal{A}_n = \sigma(X_k, \; k \leq n)$, $\mathcal{A}_\infty = \sigma(X_k, \; k \geq 0)$, $\mathcal{B}_n = \sigma(X_k, \; k \geq n)$. D'après la prop.2.2.4, $\mathcal{A}_n$ est indépendante de $\mathcal{B}_{n+1}$ et de $\mathcal{B}_\infty \subset \mathcal{B}_{n+1}$. Donc $\mathcal{B}_\infty$ est indépendante de $\mathcal{A}_\infty$ (prop.2.2.3 appliquée à $\cup \mathcal{A}_n$) mais $\mathcal{B}_\infty \subset \mathcal{A}_\infty$ d'où $\mathcal{B}_\infty$ est indépendante d'elle même. Si $A \in \mathcal{B}_\infty$, on a donc $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$ i.e. $\mathbb{P}(A) = 0$ ou 1. Si X est une v.a. $\mathcal{B}_\infty$-mesurable, $\mathbb{P}(X \leq a) = 0$ ou 1. Donc si $c = \sup(a, \mathbb{P}(X \leq a) = 0)$,

$$\mathbb{P}(X = c) = \lim \downarrow \mathbb{P}(X \leq c + \epsilon) - \lim \uparrow \mathbb{P}(X \leq c - \epsilon) = 1 \;∎$$

Applications. Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a.r. indépendantes. On a $\{\sum X_n \text{ converge}\} \in \mathcal{B}_\infty$ donc une série de v.a.r. indépendantes converge p.s. ou diverge p.s. De même $\overline{\lim} \frac{1}{n}(X_1 + \ldots + X_n)$ est une v.a. $\mathcal{B}_\infty$-mesurable et donc cette $\overline{\lim}$ est p.s. constante.

Lemme de Borel-Cantelli.

Rappelons que, pour des événements A_n,

$$\overline{\lim} \, A_n = \cap_n \cup_{k \geq n} A_k = \lim \downarrow_n \cup_{k \geq n} A_k.$$

On a donc $\overline{\lim} \, A_n = \{\omega, \; \omega \in A_n \text{ pour une infinité de } n\} = \{\sum_n \mathbb{1}_{\{A_n\}} = +\infty\}$.

Proposition 2.2.11. *Soit $(A_n, \, n \geq 0)$ une suite d'événements.*

(i) *Si* $\sum_n \mathbb{P}(A_n) < +\infty$, $\mathbb{P}(\overline{\lim}\, A_n) = 0$.

(ii) *Si les* A_n *sont indépendants et si* $\sum_n \mathbb{P}(A_n) = +\infty$, $\mathbb{P}(\overline{\lim}\, A_n) = 1$.

`Preuve:`
(i) On a

$$\mathbb{P}(\overline{\lim}\, A_n) = \lim \downarrow_n \mathbb{P}(\cup_{k\geq n} A_k) \leq \lim \downarrow_n \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0.$$

(ii) Vu l'inégalité $1 - u \leq e^{-u}$ et l'indépendance des A_n^c, on a

$$\mathbb{P}(\cap_{k=n}^m A_k^c) = \prod_{k=n}^m \mathbb{P}(A_k^c) = \prod_{k=n}^m (1 - \mathbb{P}(A_k)) \leq \exp(-\sum_{k=n}^m \mathbb{P}(A_k))$$
$$\text{donc } \mathbb{P}(\cap_{k=n}^\infty A_k^c) = \lim \downarrow_m \mathbb{P}(\cap_{k=n}^m A_k^c) = 0 \text{ si } \sum \mathbb{P}(A_n) = +\infty.$$

Passant au complémentaire, on a $\mathbb{P}(\cup_{k=n}^\infty A_k) = 1$ et $\mathbb{P}(\overline{\lim}\, A_n) = 1$ ∎

2.3. Variables aléatoires réelles

Dans la suite L^p désigne $L^p(\Omega, \mathcal{A}, \mathbb{P})$. On ne distinguera pas deux v.a.r. égales p.s. ce qui fait qu'on désigne par X aussi bien la v.a. X que sa classe d'équivalence dans L^0. En particulier on écrira pour une v.a. $X \in L^p$ aussi bien que $X \in \mathcal{L}^p$.

Moments.

Remarquons que $L^p \subset L^q$ si $p \geq q$ puisqu'alors $|X|^q \leq 1 + |X|^p$.

Définition 2.3.1. *Soit X une v.a.r. Pour $p \in [1, +\infty[$, $\mathbb{E}|X|^p$ s'appelle moment absolu d'ordre p de X; pour $p \in \mathbb{N}^*$, si $X \in L^p$, $\mathbb{E}(X^p)$ s'appelle moment d'ordre p de X.*

Notons que $\mathbb{E}|X|^p = \int |x|^p \, d\mu_X(x)$, $\mathbb{E}(X^p) = \int x^p \, d\mu_X(x)$. Les deux moments les plus importants sont le moment d'ordre 1 qui n'est rien d'autre que l'espérance de X (on dit aussi la moyenne de X) et le moment d'ordre 2. On pose, pour $X \in L^2$,

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2. \tag{2.4}$$

On a $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ et

Lemme 2.3.2. *Si $X \in L^2$, $\mathbb{E}(X - a)^2$ est minimum pour $a = \mathbb{E}(X)$ et ce minimum est* $\text{Var}(X)$.

`Preuve:`
En effet, si $m = \mathbb{E}(X)$, $\mathbb{E}(X - a)^2 = \mathbb{E}(X - m)^2 + (m - a)^2$ ∎

On note aussi σ_X^2 pour $\text{Var}(X)$, σ_X s'appelle l'écart type. Une v.a. $X \in L^1$ est dite <u>centrée</u> si $\mathbb{E}(X) = 0$. Une v.a. $X \in L^2$ est dite <u>centrée réduite</u> si $\mathbb{E}(X) = 0$ et $\mathbb{E}(X^2) = \text{Var}(X) = 1$. Noter que, si $X \in L^2$, $\sigma_X^{-1}(X - \mathbb{E}(X))$ est centrée réduite.

Proposition 2.3.3. *(i) Soit $X \in L^p$, $p \geq 1$. On a pour tout $\lambda > 0$,*

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{1}{\lambda^p}\mathbb{E}|X|^p.$$

(ii) Soit $X \in L^2$. On a $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{1}{\lambda^2}\mathrm{Var}(X).$$

`Preuve:`
(i) On remarque que $\lambda^p \mathbb{1}_{\{\{|X|\geq\lambda\}\}} \leq |X|^p$ et on prend l'espérance.

(ii) On applique (i) à $|X - \mathbb{E}(X)|$ ∎

La première de ces inégalités s'appellent l'inégalité de Markov, la seconde l'inégalité de Bienaymé-Tchebichev. Montrons maintenant l'inégalité de Jensen.

Proposition 2.3.4. *Soit X une v.a.r. et f une application convexe de $\mathbb{R}$ dans $\mathbb{R}$. On suppose X et $f(X)$ intégrables. Alors $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$.*

`Preuve:`
Soit $m = \mathbb{E}(X)$. La fonction f étant convexe, il existe une droite passant par $(m, f(m))$ et située sous le graphe de f i.e. une fonction affine $\alpha(x) = a(x-m)+f(m) \leq f(x)$ pour tout $x \in \mathbb{R}$. On a donc $a(X-m)+f(m) \leq f(X)$ et, prenant l'espérance $f(m) \leq \mathbb{E}(f(X))$
∎

Lois usuelles sur $\mathbb{R}$.

a. Loi hypergéométrique. Une urne contient n boules, n_1 blanches et n_2 noires. On en tire r sans remise. Soit X le nombre de boules blanches obtenues. On a

$$\mathbb{P}(X = x) = \frac{C_{n_1}^x C_{n_2}^{r-x}}{C_n^r}, \;\; 0 \leq x \leq n_1, \; 0 \leq r - x \leq n_2.$$

b. Loi binomiale $B(r,p)$, $0 < p < 1$, $r \in \mathbb{N}$. Une urne contient n boules, n_1 blanches et n_2 noires. On en tire r avec remise. Soit X le nombre de boules blanches obtenues. On a, posant $p = n_1/n$,

$$\mathbb{P}(X = x) = C_r^x p^x (1-p)^{r-x}, \;\; 0 \leq x \leq r.$$

On a $\mathbb{E}(X) = rp$, $\mathrm{Var}(X) = rp(1-p)$.

c. Loi de Poisson $\mathcal{P}(\lambda)$, $\lambda > 0$. C'est la loi sur $\mathbb{N}$ définie par

$$\mu(x) = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}, \;\; x = 0, 1, \ldots, n, \ldots$$

Si X est une v.a. de loi $\mathcal{P}(\lambda)$, ce qu'on note $X \sim \mathcal{P}(\lambda)$, on a $\mathbb{E}(X) = \lambda$, $\mathrm{Var}(X) = \lambda$.

d. Loi géométrique de paramètre a, $0 < a < 1$. C'est la loi sur $\mathbb{N}$ définie par

$$\mu(x) = (1-a)a^x, \;\; x = 0, 1, \ldots, n, \ldots$$

e. Loi uniforme sur $[a,b]$ notée $U(a,b)$. C'est la loi sur $\mathbb{R}$ de densité

$$h(x) = \frac{1}{b-a}\mathbb{1}_{\{[a,b]\}}(x).$$

Si $X \sim U(a,b)$, $\mathbb{E}(X) = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$.

f. Loi gamma de paramètre a, c, $a > 0$, $c > 0$, notée $G(a,c)$. Rappelons que la fonction

$$\Gamma(a) = \int_0^{+\infty} \mathrm{e}^{-x} x^{a-1}\, dx \tag{2.5}$$

est définie pour tout $a > 0$ et que l'on a $\Gamma(a+1) = a\Gamma(a)$ et $\Gamma(n) = (n-1)!$. Donc

$$h_{a,c}(x) = \frac{c^a}{\Gamma(a)}\mathrm{e}^{-cx}x^{a-1}\mathbb{1}_{\{\mathbb{R}^+\}}(x) \tag{2.6}$$

est une densité de probabilité sur $\mathbb{R}$. La loi de densité $h_{a,c}$ s'appelle la loi $G(a,c)$. On a, si $X \sim G(a,c)$, $\mathbb{E}(X) = a/c$, $\text{Var}(X) = a/c^2$.

g. Loi normale ou de Gauss $N_1(m,\sigma^2)$. On appelle loi $N_1(m,\sigma^2)$ la loi sur $\mathbb{R}$ de densité

$$f_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}. \tag{2.7}$$

Si $X \sim N_1(m,\sigma^2)$, $\mathbb{E}(X) = m$, $\text{Var}(X) = \sigma^2$. Notons que si $X \sim N_1(0,1)$, $Y = m + \sigma X \sim N_1(m,\sigma^2)$.

Fonction de répartition.

Définition 2.3.5. *Soit X une v.a.r de loi μ_X. La fonction $F_X(t) = \mathbb{P}(X \leq t) = \mu_X(]-\infty,t])$ s'appelle la fonction de répartition de X.*

La fonction F_X est croissante, continue à droite et $\lim_{t\to-\infty} F_X(t) = 0$, $\lim_{t\to+\infty} F_X(t) = 1$. La fonction F_X a donc des limites à gauche qu'on note $F_X(t-)$. On a alors

$$\mathbb{P}(X = t) = \mu_X(\{t\}) = F_X(t) - F_X(t-).$$

Si X a une densité h, on a $F_X(t) = \int_{-\infty}^t h(u)\, du$ ce qui implique $F'(t) = h(t)$ p.p.

Plus généralement on appelle fonction de répartition toute application de $\mathbb{R}$ dans $\mathbb{R}$ croissante de 0 à 1 et continue à droite. Le th.1.2.5 montre que toute fonction de répartition F est la fonction de répartition d'une probabilité sur $\mathbb{R}$ et donc qu'il existe une v.a. X de fonction de répartition F. Le théorème suivant fournit un moyen de construire une telle v.a. à partir d'une v.a. de loi $U(0,1)$.

Théorème 2.3.6. *Soient F une fonction de répartition et U une v.a.r de loi $U(0,1)$. On pose $F^{-1}(u) = \inf(t,\, F(t) \geq u)$. Alors $X = F^{-1}(U)$ a pour fonction de répartition F.*

`Preuve:`
Vu la continuité à droite de F, on a $\{t,\, F(t) \geq u\} = [F^{-1}(u), +\infty[$ et donc

$$u \leq F(t) \Leftrightarrow F^{-1}(u) \leq t.$$

Alors $\mathbb{P}(X \leq t) = \mathbb{P}(F^{-1}(U) \leq t) = \mathbb{P}(U \leq F(t)) = F(t)$ et X a pour fonction de répartition F ∎

2.4. Variables aléatoires vectorielles

Notations.

- On note, pour $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $|x| = (x_1^2 + \ldots + x_d^2)^{1/2}$.
- On note M' la matrice transposée de la matrice M. On peut représenter $x \in \mathbb{R}^d$ par un vecteur colonne i.e. une matrice $d \times 1$ et on écrira indifferemment $x = (x_1, \ldots, x_d)$ ou $x = (x_1 \ldots x_d)'$. Pour $x = (x_1 \ldots x_d)'$ et $y = (y_1 \ldots y_d)'$, on a $x'y = x_1y_1 + \ldots + x_dy_d =< x, y >$ et xy' est la matrice de terme général x_iy_j.
- On note $L_d^p = \{X = (X_1, \ldots, X_d),\ \mathbb{E}|X|^p < +\infty\}$.
- Si $X \in L_d^1$, on note $\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d))$.

$X = (X_1, \ldots, X_d)'$ est un vecteur aléatoire ssi, pour $k = 1, \ldots, d$, X_k est une v.a.r. Ceci entraîne:

Proposition 2.4.1. *Soient $X_1, \ldots, X_n$ des v.a.r. et $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Une v.a.r. Y est $\mathcal{F}_n$-mesurable ssi $Y = f(X_1, \ldots, X_n)$ avec f fonction borélienne sur $\mathbb{R}^d$.*

`Preuve:`
Soit $X = (X_1, \ldots, X_n)$. On a $\mathcal{F}_n = \sigma(X)$ et il suffit d'appliquer la prop.1.1.4 ∎

Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire. Les lois $\mu_{X_1}, \ldots, \mu_{X_d}$ s'appellent les lois marginales de X. On sait (th.2.2.7) que les composantes $X_1, \ldots, X_d$ sont indépendantes ssi $\mu_X = \mu_{X_1} \otimes \ldots \otimes \mu_{X_d}$. Si X a une densité h, on a immédiatement:

Lemme 2.4.2. *Soit X un vecteur aléatoire de densité h. Les composantes $X_1, \ldots, X_d$ sont indépendantes ssi $h(x_1, \ldots, x_d) = g_1(x_1) \ldots g_d(x_d)$ p.p. et alors X_k a pour densité $h_{X_k}(u) = g_k(u)/\int_{\mathbb{R}} g_k(v)\, dv$.*

Rappelons la formule de changement de variables dans $\mathbb{R}^d$. Si ϕ est un difféomorphisme de l'ouvert U sur l'ouvert V, on a, pour toute $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\int_V f(v)\, dv = \int_U f(\phi(u))|J(\phi)(u)|\, du. \tag{2.8}$$

où $J(\phi)$ est le déterminant de la matrice des $\frac{\partial \phi_j}{\partial u_k}$. Il en résulte:

Proposition 2.4.3. *Soit X un vecteur aléatoire de densité h. On suppose que $X \in D$ p.s., D ouvert de $\mathbb{R}^d$. Soient ψ un difféomorphisme de D sur un ouvert Δ et $Y = \psi(X)$, alors Y a pour densité*

$$h(\psi^{-1}(y))|J(\psi^{-1})(y)|\mathbb{1}_{\{\Delta\}}(y).$$

`Preuve:`
On a, pour toute $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\mathbb{E}(f(Y)) = \mathbb{E}(f(\psi(X))) = \int_D f(\psi(x))h(x)\,dx = \int_\Delta f(y)h(\psi^{-1}(y)|J(\psi^{-1})(y)|\,dy \;\blacksquare$$

Exemple. Soient X et Y des v.a.r. indépendantes de lois respectives $G(a,c)$ et $G(b,c)$ (2.6), $a,b,c > 0$. On pose $S = X+Y$, $T = \frac{X}{X+Y}$. On veut calculer la loi du couple (S,T). Vu l'indépendance, le couple (X,Y) a pour densité

$$h_{X,Y}(x,y) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)}\mathrm{e}^{-c(x+y)}x^{a-1}y^{b-1}\mathbb{1}_{\{]0,+\infty[\}}(x)\mathbb{1}_{\{]0,+\infty[\}}(y).$$

Soit ϕ l'application $(x,y) \mapsto (s = x+y,\, t = \frac{x}{x+y})$. ϕ est un difféomorphisme de $]0,+\infty[\times]0,+\infty[$ sur $]0,+\infty[\times]0,1[$. De plus $J(\phi^{-1})(s,t) = -s$. La densité de (S,T) est donc (prop.2.4.3)

$$h_{S,T}(s,t) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)}\mathrm{e}^{-cs}s^{a+b-1}t^{a-1}(1-t)^{b-1}\mathbb{1}_{\{]0,+\infty[\}}(s)\mathbb{1}_{\{]0,1[\}}(t).$$

Le lem.2.4.2 montre que S et T sont indépendantes, que S a pour densité

$$h_S(s) = \frac{c^{a+b}}{\Gamma(a+b)}\,\mathrm{e}^{-cs}s^{a+b-1}\mathbb{1}_{\{]0,+\infty[\}}(s)$$

i.e. $S \sim G(a+b,c)$ et que T a pour densité

$$h_T(t) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,t^{a-1}(1-t)^{b-1}\mathbb{1}_{\{]0,1[\}}(t).$$

Puisque h_T est une densité de probabilité, on a montré au passage la formule

$$\int_0^1 t^{a-1}(1-t)^{b-1}\,dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Covariance.

Soient U et V deux v.a.r. de carré intégrable. On appelle covariance de U et V la quantité

$$\mathrm{Cov}(U,V) = \mathbb{E}[(U-\mathbb{E}(U))(V-\mathbb{E}(V))] = \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V).$$

$(U, V) \mapsto \text{Cov}(U, V)$ est une forme bilinéaire et $\text{Cov}(U, U) = \text{Var}(U)$. Si U et V sont indépendantes, $\text{Cov}(U, V) = 0$. On appelle coefficient de corrélation de U et V la quantité

$$\rho_{U,V} = \frac{\text{Cov}(U, V)}{\sigma_U \sigma_V}.$$

On a $|\rho_{U,V}| \leq 1$ et $|\rho_{U,V}| = 1$ ssi $aU + bV + c = 0$ p.s.
Plus généralement on pose, pour $X \in L_d^2$

$$K(X) = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))'] = \mathbb{E}(XX') - \mathbb{E}(X)[\mathbb{E}(X)]'. \tag{2.9}$$

$K(X)$ s'appelle la matrice de covariance de X. On a $K_{ij}(X) = \text{Cov}(X_i, X_j)$. Notons que, si les composantes $X_1, \ldots, X_d$ sont indépendantes, $K(X)$ est diagonale.

Proposition 2.4.4. *Soit $X \in L_d^2$. On a*

(i) $K(\alpha X) = \alpha^2 K(X)$, $\alpha \in \mathbb{R}$; $K(X + a) = K(X)$, $a \in \mathbb{R}^d$; $K'(X) = K(X)$.

(ii) *Pour tout* $\lambda \in \mathbb{R}^d$, $\lambda' K(X) \lambda \geq 0$.

(iii) *Soit M une matrice déterministe $r \times d$, on a* $K(MX) = MK(X)M'$.

`Preuve:`
(i) résulte de la définition (2.9) ∎

(ii) Vu (i), on peut supposer $\mathbb{E}(X) = 0$. Alors

$$\lambda' K(X) \lambda = \lambda' \mathbb{E}(XX') \lambda = \mathbb{E}(\lambda' XX' \lambda) = \mathbb{E}|\lambda' X|^2 \geq 0.$$

(iii) Vu (i), on peut supposer $\mathbb{E}(X) = 0$. Alors

$$K(MX) = \mathbb{E}(MX(MX)') = \mathbb{E}(MXX'M') = M\mathbb{E}(XX')M' = MK(X)M' \quad \blacksquare$$

Les points (i) et (ii) montrent que $K(X)$ est symétrique semi-définie positive.

Théorème 2.4.5. *Soient $X, Y \in L_d^2$ des vecteurs aléatoires indépendants, on a $K(X + Y) = K(X) + K(Y)$. En particulier, si $d = 1$, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ si les v.a.r. X et Y sont indépendantes.*

`Preuve:`
On peut supposer $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. Alors $K(X + Y) = \mathbb{E}((X + Y)(X + Y)') = \mathbb{E}(XX') + \mathbb{E}(YY')$ puisque, vu l'indépendance, $\mathbb{E}(XY') = \mathbb{E}(X)\mathbb{E}(Y') = 0$ et de même $\mathbb{E}(YX') = 0$ ∎

Proposition 2.4.6. *Soit $X \in L_d^2$. On a* $\mathbb{P}(X - \mathbb{E}(X) \in \text{Im}\, K(X)) = 1$.

`Preuve:`
Comme toujours on peut supposer $\mathbb{E}(X) = 0$. Soit $V = \text{Im}\, K(X)$. Si $\dim(V) = d$, il n'y a rien à montrer. Supposons $\dim(V) = r < d$. Il existe $a_1, \ldots, a_{d-r} \in \text{Ker}(X)$ tels que $x \in V$ ssi $a_k' x = 0$, $k = 1, \ldots, d - r$ (pour voir cela il suffit de se placer dans une base où $K(X)$ est diagonale). On a alors, vu la prop.2.4.4,

$$\mathbb{E}(a_k' X)^2 = \text{Var}(a_k' X) = K(a_k' X) = a_k' K(X) a_k = 0$$

d'où $a_k' X = 0$ p.s. et $X \in V$ p.s. ∎

2.5. Convergence des suites de variables aléatoires

Le chapitre 5 sur les martingales étant en grande partie consacré à ce sujet, on se contente de présenter les différents types de convergence et d'énoncer la loi des grands nombres. La convergence en loi, qui n'est que très peu utilisée dans la suite, est rejetée en annexe. Pour $x \in \mathbb{R}^d$, $|x|$ désigne une norme quelconque, les définitions ci-dessous étant indépendantes du choix de cette norme.

Les différents modes de convergence.

Définition 2.5.1. *Soient X_n et X des v.a. à valeurs $\mathbb{R}^d$.*

- *On dit que la suite X_n converge en probabilité vers la v.a. X si, pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to_n 0$.*
- *On dit que la suite X_n converge presque sûrement (en abrégé p.s.) vers la v.a. X si $\mathbb{P}\{\omega \,;\, X_n(\omega) \to_n X(\omega)\} = 1$.*
- *On dit que la suite X_n converge vers la v.a. X dans L^p, $1 \leq p < +\infty$, si $X_n, X \in L^p$ et si $\mathbb{E}|X_n - X|^p \to_n 0$.*

On vérifie immédiatement que $X_n = (X_{n,1}, \ldots, X_{n,d})$ converge vers $X = (X_1 \ldots, X_d)$ en un des sens ci-dessus ssi, pour $k = 1, \ldots, d$, $X_{n,k}$ converge vers X_k dans le même sens. On ne considérera donc plus que des v.a. réelles. De plus ces convergences sont compatibles avec la structure d'espace vectoriel, c'est à dire que si deux suites convergent, il en est de même de toute combinaison linéaire et la limite est la combinaison linéaire des limites.

On note, X étant une v.a.r., $||X||_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}$. Vu l'inégalité de Hölder (1.6), on a, pour $1 \leq p \leq q$, $||X||_p \leq ||X||_q$ et donc la convergence dans L^q implique la convergence dans L^p. En particulier la convergence dans L^2 implique la convergence dans L^1. La convergence dans L^1 s'appelle aussi la convergence en moyenne, la convergence dans L^2 s'appelle aussi la convergence en moyenne quadratique.

La définition de la convergence p.s. signifie qu'il existe un ensemble $\widetilde{\Omega} \subset \Omega$ de probabilité 1 vérifiant:

$$\forall \omega \in \widetilde{\Omega},\ \forall \varepsilon > 0,\ \exists N \text{ tel que } n \geq N \Rightarrow |X_n(\omega) - X(\omega)| \leq \varepsilon$$

ou encore, en posant $\Omega_\varepsilon = \cup_{N\geq 0} \cap_{n \geq N} \{|X_n - X| \leq \varepsilon\}$ et $\widetilde{\Omega} = \cap_{\varepsilon > 0} \Omega_\varepsilon$, on devrait avoir $\mathbb{P}(\widetilde{\Omega}) = 1$. Le problème est que, pour définir $\widetilde{\Omega}$, on utilise une intersection non dénombrable et qu'il n'est donc pas clair que cet ensemble soit mesurable. Fort heureusement, c'est bien le cas car l'application $\varepsilon \hookrightarrow \Omega_\varepsilon$ est croissante et donc $\widetilde{\Omega} = \cap_p \Omega_{1/p}$. Toute intersection dénombrable d'ensemble de probabilité 1 étant de probabilité 1, la convergence p.s. est donc équivalente à:

$$\forall \varepsilon > 0, \qquad \mathbb{P}\left(\cup_{N \geq 0} \cap_{n \geq N} \{|X_n - X| \leq \varepsilon\}\right) = 1$$

Autrement dit:

Proposition 2.5.2. *La suite X_n converge presque sûrement (en abrégé p.s.) vers la v.a. X ssi pour tout $\varepsilon > 0$, $\mathbb{P}\{\cup_{k\geq n}(|X_k - X| > \varepsilon)\} \to_n 0$.*

Proposition 2.5.3. *La convergence dans L^1 implique la convergence en probabilité, la convergence p.s. implique la convergence en probabilité.*

`Preuve:`
(i) D'après l'inégalité de Markov (prop.2.3.3), $\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-1}\mathbb{E}|X_n - X|$ ce qui montre le premier point.

(ii) C'est une conséquence immédiate de la prop.2.5.2 ∎

Notons que si X_n converge en probabilité vers X et vers Y, on a $\mathbb{P}(|X - Y| > \varepsilon) \leq \mathbb{P}(|X - X_n| > \frac{\varepsilon}{2}) + \mathbb{P}(|X_n - Y| > \frac{\varepsilon}{2}) \to_n 0$ et donc $\mathbb{P}(|X - Y| > 0) = 0$ et $X = Y$ p.s. Ceci implique, vu la prop.2.5.3, que les limites de X_n en les différents sens sont p.s. uniques et égales.

Exemple. Soit X_n une suite de v.a.r. indépendantes telles que $\mathbb{P}(X_n = a_n) = p_n$, $\mathbb{P}(X_n = 0) = 1 - p_n$. On suppose $0 < p_n < 1$, $p_n \to_n 0$ et $a_n \geq 1$.

a. On a, pour $\varepsilon \in]0, 1[$, $\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n > \varepsilon) = p_n$ et $X_n \to_n 0$ en probabilité.

b. On a $\sum \mathbb{P}(X_n > 0) = \sum p_n$ donc, si $\sum p_n < +\infty$, on a (prop.2.2.11) que $\{X_n > 0\}$ n'a p.s. lieu que pour un nombre fini de n donc $X_n \to_n 0$ p.s. Réciproquement si $\sum p_n = +\infty$, on a (prop.2.2.11) que $\{X_n = a_n\}$ a p.s. lieu pour une infinité de n donc X_n ne converge pas p.s. vers 0. Donc $X_n \to_n 0$ p.s. ssi $\sum p_n < +\infty$.

c. $\mathbb{E}|X_n| = \mathbb{E}(X_n) = a_n p_n$. Donc $X_n \to_n 0$ dans L^1 ssi $a_n p_n \to_n 0$.

d. $\mathbb{E}(X_n^2) = a_n^2 p_n$. Donc $X_n \to_n 0$ dans L^2 ssi $a_n^2 p_n \to_n 0$.

Si on choisit $p_n = \frac{1}{n}$, $a_n = 1$, X_n converge vers 0 dans L^1 mais pas p.s. Si on choisit $p_n = \frac{1}{n^2}$, $a_n = n^2$, X_n converge vers 0 p.s. mais pas dans L^1. Si on choisit $p_n = \frac{1}{n^2}$, $a_n = n$, X_n converge vers 0 dans L^1 mais pas dans L^2.

Critères de convergence p.s.

Proposition 2.5.4. *Si, pour tout $\varepsilon > 0$, $\sum \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$, alors la suite X_n converge p.s. vers X.*

`Preuve:`

$$\mathbb{P}\{\cup_{k\geq n}(|X_k - X| > \varepsilon)\} \leq \sum_{k=n}^{\infty} \mathbb{P}\{(|X_k - X| > \varepsilon)\}$$

Il suffit alors d'utiliser la prop.2.5.2 ∎

On en déduit les corollaires suivants:

Corollaire 2.5.5. *S'il existe une suite $\varepsilon_n > 0$, avec $\lim \varepsilon_n = 0$ et $\sum \mathbb{P}(|X_n - X| > \varepsilon_n) < +\infty$, alors la suite X_n converge p.s. vers X.*

Corollaire 2.5.6. *De toute suite X_n convergeant en probabilité, on peut extraire une sous-suite X_{n_k} convergeant p.s.*

Preuve:
Vu que, pour tout k, $\mathbb{P}(|X_n - X| > 2^{-(k+1)}) \to_n 0$, on peut construire une suite croissante n_k telle que $\mathbb{P}(|X_{n_k} - X| > 2^{-(k+1)}) \leq 2^{-(k+1)}$. Il suffit alors d'utiliser le corollaire 2.5.5 ∎

Les deux résultats suivants s'appuient sur la propriété de Cauchy des suites convergentes dans $\mathbb{R}$:

Proposition 2.5.7. *Soit X_n une suite de v.a.r. Si $\sum \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty$ pour une suite $\varepsilon_n > 0$ vérifiant $\sum \varepsilon_n < +\infty$, la suite X_n converge p.s.*

Preuve:
D'après le lemme de Borel-Cantelli (prop.2.2.11), il existe un ensemble $\widetilde{\Omega}$ de probabilité 1 tel que pour tout $\omega \in \widetilde{\Omega}$, il existe $n_0(\omega)$ tel que, pour tout $n \geq n_0(\omega)$, $|X_{n+1}(\omega) - X_n(\omega)| \leq \varepsilon_n$. On a donc, pour $n > m \geq n_0(\omega)$,

$$|X_n(\omega) - X_m(\omega)| \leq \sum_{k=m}^{n-1} |X_{k+1}(\omega) - X_k(\omega)| \leq \sum_{k=m}^{n-1} \varepsilon_k.$$

Vu la convergence de $\sum \varepsilon_n$, ceci implique que $X_n(\omega)$ est une suite de Cauchy et donc $X_n(\omega)$ converge ∎

Proposition 2.5.8. *Une suite X_n de v.a.r. converge p.s. ssi, pour tout $\epsilon > 0$,*

$$\lim_{N\to\infty} \mathbb{P}(\sup_{p\geq 0} |X_{p+N} - X_N| > \epsilon) = 0$$

Preuve:
Une suite converge p.s. si et seulement si elle est de Cauchy pour presque tout ω. Autrement dit, il existe un ensemble $\widetilde{\Omega} \subset \Omega$ de probabilité 1 vérifiant:

$$\forall \omega \in \widetilde{\Omega},\ \forall \varepsilon > 0,\ \exists N \text{ tel que } p, q \geq N \Rightarrow |X_p(\omega) - X_q(\omega)| \leq \varepsilon$$

ou encore, en posant $\Omega_\varepsilon = \cup_{N\geq 0} \cap_{(p,q)\geq 0} \{|X_{N+p} - X_{N+q}| \leq \varepsilon\}$ et $\widetilde{\Omega} = \cap_{\varepsilon > 0} \Omega_\varepsilon$, on devrait avoir $\mathbb{P}(\widetilde{\Omega}) = 1$. Le problème est que, pour définir $\widetilde{\Omega}$, on utilise de nouveau une intersection non dénombrable et qu'il n'est donc pas clair que cet ensemble soit mesurable. Comme dans le cas précédent, il suffit de remarquer que l'application $\varepsilon \hookrightarrow \Omega_\varepsilon$ est croissante et donc $\widetilde{\Omega} = \cap_p \Omega_{1/p}$. Toute intersection dénombrable d'ensemble de probabilité 1 étant de probabilité 1, la convergence p.s. est donc équivalente à:

$$\forall \varepsilon > 0, \qquad \mathbb{P}\left(\cup_{N\geq 0} \cap_{(p,q)\geq 0} \{|X_{N+p} - X_{N+q}| \leq \varepsilon\}\right) = 1$$

Soit $\epsilon > 0$. On pose $A_N = \cup_{(p,q)\geq 0}(|X_{p+N} - X_{N+q}| > \epsilon))$, on a alors

$$A_N \subset (\cup_{p\geq 0}(|X_{p+N} - X_N| > \epsilon/2)) \cup (\cup_{q\geq 0}(|X_{q+N} - X_N| > \epsilon/2))$$

La propriété de l'énoncé implique que $\lim_N \mathbb{P}(A_N) = 0$ et la suite A_N étant décroissante, on obtient que l'ensemble $\Omega_\epsilon = \cup_N(\cap_{p,q\geq 0}(|X_{p+N} - X_{N+q}| \leq \epsilon))$ est de probabilité 1. ∎

2.6. Intégrabilité uniforme.

Soit X une v.a.r. intégrable. On a $|X| < +\infty$ p.s. et donc $|X|\mathbb{1}_{\{|X|>a\}} \to 0$ p.s. lorsque $a \to +\infty$. Puisque $|X|\mathbb{1}_{\{|X|>a\}} \leq |X| \in L^1$, on a (th. de Lebesgue) $\mathbb{E}(|X|\mathbb{1}_{\{|X|>a\}}) \to_{a\to+\infty} 0$. Ceci est à la base de la notion d'uniforme intégrabilité (on dit aussi équi-intégrabilité).

Définition 2.6.1. *On dit qu'une famille $\mathcal{H}$ de v.a. réelles est uniformément intégrable si*

$$\sup_{X\in\mathcal{H}} \int_{\{|X|>a\}} |X|\, d\mathbb{P} \to_{a\to+\infty} 0.$$

Ceci équivaut à:

$$\text{pour tout}\, \varepsilon > 0, \text{ il existe } a_0 \text{ tel que, pour tous } X \in \mathcal{H} \text{ et } a \geq a_0,\ \int_{\{|X|>a\}} |X|\, d\mathbb{P} \leq \varepsilon.$$

Exemple: Soit $Z \in L^1$, $Z \geq 0$. La famille $\mathcal{H} = \{X,\ |X| \leq Z\}$ est uniformément intégrable puisque, si $|X| \leq Z$, $\int_{\{|X|>a\}} |X|\, d\mathbb{P} \leq \int_{\{Z>a\}} Z\, d\mathbb{P}$. Ceci implique qu'une famille finie de v.a.r. intégrables est uniformément intégrable.

Un outil de base sera:

Proposition 2.6.2. *La famille $\mathcal{H}$ est uniformément intégrable ssi*

(i) *La famille $\mathcal{H}$ est bornée dans L^1, i.e.* $\sup_{X\in\mathcal{H}} \mathbb{E}|X| < +\infty$,

(ii) *La famille $\mathcal{H}$ est équicontinue, i.e., pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que $\mathbb{P}(A) < \alpha$ implique* $\sup_{X\in\mathcal{H}} \int_A |X|\, d\mathbb{P} < \varepsilon$.

`Preuve:`
a. Supposons la famille $\mathcal{H}$ uniformément intégrable. On a, pour tous $A \in \mathcal{A}$ et $X \in \mathcal{H}$,

$$\int_A |X|\, d\mathbb{P} = \int_A |X|\mathbb{1}_{\{|X|\leq a\}}\, d\mathbb{P} + \int_A |X|\mathbb{1}_{\{|X|>a\}}\, d\mathbb{P} \leq a\mathbb{P}(A) + \int |X|\mathbb{1}_{\{|X|>a\}}\, d\mathbb{P}.$$

Pour $A = \Omega$, on a $\sup_{X\in\mathcal{H}} \mathbb{E}|X| < +\infty$. Pour montrer (ii), on choisit a tel que, pour toute $X \in \mathcal{H}$, $\sup_{X\in\mathcal{H}} \mathbb{E}(|X|\mathbb{1}_{\{|X|>a\}}) < \frac{\varepsilon}{2}$ et on a, pour tout $A \in \mathcal{H}$, $\int_A |X|\, d\mathbb{P} < \varepsilon$ si $\mathbb{P}(A) < \alpha = \frac{\varepsilon}{2a}$.

b. Supposons (i) et (ii). Soit $M = \sup_{X\in\mathcal{H}} \mathbb{E}|X|$. On a $\mathbb{P}(|X| > a) \leq \frac{1}{a}\mathbb{E}|X|$ et donc, pour toute $X \in \mathcal{H}$, $\mathbb{P}(|X| > a) \leq \frac{M}{a}$ et il suffit d'utiliser (ii) ∎

La prop.2.6.2 va nous permettre d'établir:

Proposition 2.6.3. *Soit $1 \leq p < \infty$. Si X_n converge dans L^p, la famille $|X_n|^p_{n\geq 0}$ est uniformément intégrable.*

`Preuve:`
On vérifie (i) et (ii) de la prop.2.6.2. Soit X la limite de X_n dans L^p. Vu l'inégalité $|a+b|^p \leq 2^{p-1}(|a|^p + |b|^p)$ on obtient que $\mathbb{E}|X_n|^p \leq 2^{p-1}(\mathbb{E}|X - X_n|^p + \mathbb{E}|X|^p)$, et donc

$\sup_n \mathbb{E}|X_n|^p < +\infty$. Montrons (ii). Soit $\varepsilon > 0$. Il existe $\alpha_1 > 0$ tel que $\int_A |X|^p \, d\mathbb{P} < \varepsilon/2$ si $\mathbb{P}(A) < \alpha_1$. Vu que

$$\int_A |X_n|^p \, d\mathbb{P} \le 2^{p-1}\Big(\int_A |X|^p \, d\mathbb{P} + \int_A |X_n - X|^p \, d\mathbb{P}\Big) \le 2^{p-1}\Big(\int_A |X|^p \, d\mathbb{P} + \mathbb{E}|X_n - X|^p\Big),$$

il existe n_0 tel que $\int_A |X_n|^p \, d\mathbb{P} < \varepsilon$ si $n \ge n_0$ et $\mathbb{P}(A) < \alpha_1$. La famille $(|X_1|^p \ldots, |X_{n_0}|^p)$ étant uniformément intégrable, il existe $\alpha_2 > 0$ tel que $\int_A |X_k|^p \, d\mathbb{P} < \varepsilon$ si $\mathbb{P}(A) < \alpha_2$ et $k \le n_0$. On a établi (ii) avec $\alpha = \alpha_1 \wedge \alpha_2$ ∎

Donnons un critère très pratique d'uniforme intégrabilité.

Proposition 2.6.4. *Soit g une application de $\mathbf{R}^+$ dans $\mathbf{R}^+$ telle que $\lim_{t \to +\infty} \frac{g(t)}{t} = +\infty$. Si $\sup_{X \in \mathcal{H}} \mathbb{E}[g(|X|)] < +\infty$, la famille $\mathcal{H}$ est uniformément intégrable.*

`Preuve:`
Soient $M = \sup_{X \in \mathcal{H}} E[g(|X|)]$ et $\varepsilon > 0$. Par hypothèse, il existe A tel que, si $t > A$, $\frac{g(t)}{t} > \frac{M}{\varepsilon}$. On a alors, pour tout $a > A$ et tout $X \in \mathcal{H}$,

$$\int_{\{|X|>a\}} |X| \, d\mathbb{P} \le \frac{\varepsilon}{M} \int_{\{|X|>a\}} g(|X|) \, d\mathbb{P} \le \frac{\varepsilon}{M} \mathbb{E}[g(|X|)] \le \varepsilon$$ ∎

Exemple: On choisit $g(t) = t^p$, $p > 1$. Alors si $\sup_{X \in \mathcal{H}} ||X||_p < +\infty$, la famille $\mathcal{H}$ est uniformément intégrable ∎

Le résultat essentiel est:

Théorème 2.6.5. *Soit X_n une suite de v.a. réelles convergeant en probabilité. Alors X_n converge dans L^p ssi la famille $|X_n|^p_{n \ge 0}$ est uniformément intégrable.*

`Preuve:`
Supposons que la famille $|X_n|^p_{n \ge 0}$ soit uniformément intégrable et que $X_n \to X$ en probabilité. On a (prop.2.6.2) $\sup_n \mathbb{E}|X_n|^p < +\infty$. D'après le cor.2.5.6, il existe une sous-suite X_{n_k} convergeant p.s. vers X et, vu le lemme de Fatou, $\mathbb{E}|X|^p \le \underline{\lim}_k \mathbb{E}|X_{n_k}|^p \le \sup_n \mathbb{E}|X_n|^p < +\infty$. Donc $X \in L^p$. On a alors, pour tout $1 > \varepsilon > 0$,

$$\begin{aligned} \mathbb{E}|X_n - X|^p &= \mathbb{E}(|X_n - X|^p \mathbb{1}_{\{|X_n - X| \le \varepsilon\}}) + \mathbb{E}(|X_n - X|^p \mathbb{1}_{\{|X_n - X| > \varepsilon\}}) \\ &\le \varepsilon^p + 2^{p-1}(\mathbb{E}(|X_n|^p \mathbb{1}_{\{|X_n - X| > \varepsilon\}}) + \mathbb{E}(|X|^p \mathbb{1}_{\{|X_n - X| > \varepsilon\}})). \end{aligned}$$

On en déduit alors facilement, utilisant la prop.2.6.2, que $\mathbb{E}|X_n - X|^p \le 2\varepsilon$ si n est assez grand. Donc $X_n \to_n X$ dans L^p. La réciproque résulte de la prop.2.6.3 ∎

Remarque: Le th.2.6.5 est une généralisation du théorème de Lebesgue puisque pour $Y \in L^1$, $\{X_n, \; |X_n| \le Y\}$ est uniformément intégrable.

Loi des grands nombres.

Proposition 2.6.6. *Soit X_n une suite de v.a.r. de carré intégrable. On suppose que $\sup_n \mathrm{Var}(X_n) = K < +\infty$, que $Cov(X_i, X_j) = 0$, $i \neq j$ et que $\frac{1}{n}(\mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n)) \to_n m$. Alors*

$$\frac{1}{n}(X_1 + \ldots + X_n) \to_n m \ \textit{ p.s. et dans } L^2.$$

`Preuve:`
On pose $S_n = X_1 + \ldots + X_n$. Puisque $S_n = \sum_{k=1}^n (X_k - \mathbb{E}(X_k)) + \sum_{k=1}^n \mathbb{E}(X_k)$, on peut supposer $\mathbb{E}(X_k) \equiv 0$ ce qu'on fait dorénavant.

(i) Puisque $\mathbb{E}(X_i X_j) = 0$, $i \neq j$, et $m = 0$, on a $\mathbb{E}(S_n^2) = \sum_{k=1}^n \mathbb{E}(X_k^2) \leq Kn$. Donc $\mathbb{E}(\frac{S_n}{n})^2 \to_n 0$ i.e. $\frac{S_n}{n} \to_n 0$ dans L^2.

(ii) On a $\mathbb{P}(|S_{n^2}|/n^2 > \varepsilon) = \mathbb{P}(|S_{n^2}| > \varepsilon n^2) \leq \mathbb{E}(S_{n^2}^2)/\varepsilon^2 n^4 \leq K/\varepsilon^2 n^2$. Donc (lem.2.5.4) $S_{n^2}/n^2 \to_n 0$ p.s.

(iii) On pose $W_n = \sup(|S_k - S_{n^2}|, \ n^2 + 1 \leq k \leq (n+1)^2)$. On a $\mathbb{P}(W_n/n^2 > \varepsilon) \leq \mathbb{E}(W_n^2)/\varepsilon^2 n^4$. Mais $W_n^2 \leq \sum_{k=n^2+1}^{(n+1)^2} (S_k - S_{n^2})^2$ et $\mathbb{E}(W_n^2) \leq K(2n+1)^2$. Donc $\mathbb{P}(W_n/n^2 > \varepsilon) \leq K(2n+1)^2/\varepsilon^2 n^4$ et ((lem.2.5.4) $W_n/n^2 \to_n 0$ p.s.

(iv) Notons m_n la partie entière de $\sqrt{n}$. On a $m_n \leq \sqrt{n} \leq m_n + 1$ et

$$|\frac{S_n}{n} - \frac{S_{m_n^2}}{m_n^2}| \leq \frac{|S_n - S_{m_n^2}|}{n} + |S_{m_n^2}|(\frac{1}{m_n^2} - \frac{1}{n}) \leq \frac{W_{m_n}}{m_n^2} + \frac{|S_{m_n^2}|}{m_n^2}(1 - \frac{m_n^2}{n}).$$

Par définition $\frac{m_n^2}{n} \to_n 1$, vu (iii), $\frac{W_{m_n}}{m_n^2} \to_n 0$ p.s. et, vu (ii), $\frac{S_{m_n^2}}{m_n^2} \to_n 0$ p.s. Ceci montre que $\frac{S_n}{n} \to_n 0$ p.s. ∎

Considérons maintenant une suite X_n de v.a.r. indépendantes de même loi avec $\mathbb{E}(X_1^2) < +\infty$. On peut appliquer la prop.2.6.6 et l'on a $\frac{S_n}{n} \to_n \mathbb{E}(X_1)$ p.s. C'est la loi forte des grands nombres. En fait il suffit d'avoir $\mathbb{E}|X_1| < +\infty$ comme on verra au chapitre 5 et l'on a:

Théorème 2.6.7. *Soit X_n une suite de v.a.r. indépendantes de même loi avec $\mathbb{E}|X_1| < +\infty$. On a*

$$\frac{1}{n}(X_1 + \ldots + X_n) \to_n \mathbb{E}(X_1) \ \textit{ p.s. et dans } L^1.$$

2.7. Annexe

2.7.1. Fonctions caractéristiques

La plupart des résultats de cette section sont une simple traduction en termes probabilistes de ceux de la section 1.6.

Soient X et Y deux v.a. indépendantes à valeurs $\mathbb{R}^d$. On pose $S = X + Y$. Cherchons la loi de S. On a, pour toute $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\mathbb{E}(f(S)) = \mathbb{E}(f(X+Y)) = \int f(x+y)\, d\mu_X(x) d\mu_Y(y) = \int f\, d(\mu_X * \mu_Y)$$

d'après (1.9). On a donc

Proposition 2.7.1. *Soient X et Y deux v.a. indépendantes à valeurs* $\mathbb{R}^d$*. On a* $\mu_{X+Y} = \mu_X * \mu_Y$.

Définition 2.7.2. *Soit X une v.a. à valeurs* $\mathbb{R}^d$*. On appelle fonction caractéristique de X la fonction sur* $\mathbb{R}^d$

$$\phi_X(t) := \hat{\mu}_X(t) = \mathbb{E}(e^{i<t,X>}).$$

Propriétés élémentaires.

- ϕ_X est continue, $\phi_X(0) = 1$, $|\phi_X(t)| \leq 1$.
- $\phi_{\alpha X+b}(t) = \mathrm{e}^{i<t,b>}\phi_X(\alpha t)$, $\alpha \in \mathbb{R}$, $b \in \mathbb{R}^d$; $\phi_{-X}(t) = \overline{\phi_X(t)}$; si $\mu_{-X} = \mu_X$, ϕ_X est réelle.

Du th.1.6.8 résulte:

Théorème 2.7.3. *Soient X et Y deux v.a. à valeurs* $\mathbb{R}^d$*. Si* $\phi_X = \phi_Y$*,* $\mu_X = \mu_Y$*. Si* $\phi_X \in L^1(\lambda)$*,* $\mu_X = h.\lambda$ *avec*

$$h(x) = (2\pi)^{-d} \int e^{-i<t,x>}\phi_X(t)\, dt.$$

De même le th.1.6.3 et la prop.2.7.1 impliquent

Théorème 2.7.4. *Soient X et Y deux v.a. indépendantes à valeurs* $\mathbb{R}^d$*. On a* $\phi_{X+Y} = \phi_X\phi_Y$.

`Preuve:`
En fait cela se montre immédiatement grâce au cor.2.2.9

$$\phi_{X+Y}(t) = \mathbb{E}(\mathrm{e}^{i<t,X+Y>}) = \mathbb{E}(\mathrm{e}^{i<t,X>}\mathrm{e}^{i<t,Y>}) = \mathbb{E}(\mathrm{e}^{i<t,X>})\mathbb{E}(\mathrm{e}^{i<t,Y>}) = \phi_X(t)\phi_Y(t).$$

Le théorème suivant est souvent appelé “critère Fourier d'indépendance”.

Théorème 2.7.5. *Des v.a.* $X_1,\ldots,X_n$ *à valeurs* $\mathbb{R}^{d_1},\ldots,\mathbb{R}^{d_n}$ *sont indépendantes ssi, pour tous* $t_1 \in \mathbb{R}^{d_1},\ldots,$ *tout* $t_n \in \mathbb{R}^{d_n}$,

$$\phi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n) = \phi_{X_1}(t_1)\ldots\phi_{X_n}(t_n).$$

`Preuve:`
En effet cette condition est équivalente (th.2.7.3) à $\mu_{(X_1,\ldots,X_n)} = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$ ce qui équivaut (th.2.2.7) à l'indépendance de $X_1,\ldots,X_n$ ∎

Proposition 2.7.6. *Soit X une v.a. à valeurs* $\mathbb{R}^d$.

(i) *Si* $X \in L^1_d$*,* ϕ_X *est dérivable et* $\frac{\partial \phi_X}{\partial t_k}(0) = i\mathbb{E}(X_k)$.

(ii) *Si $X \in L^2_d$, ϕ_X est deux fois dérivable et $\frac{\partial^2 \phi_X}{\partial t_j \partial t_k}(0) = -\mathbb{E}(X_j X_k)$.*

`Preuve:`
(i) On remarque que $|\frac{\partial}{\partial t_k} \mathrm{e}^{i<t,X>}| = |X_k| \in L^1$ et on applique la prop.1.3.7.

(ii) On continue... ∎

Il est facile de voir en appliquant la prop.1.3.7 que si $X \in L^m_d$, ϕ_X est m fois dérivable. Réciproquement on a ,

Proposition 2.7.7. *Soit X une v.a. à valeurs $\mathbb{R}^d$. Si ϕ_X est $2m$ fois dérivable en 0, m entier, $X \in L^{2m}_d$.*

`Preuve:`
On se limite à $d = 1$, $m = 1$. On pose $\phi = \phi_X$ et $\mu = \mu_X$. On a $\phi"(0) = \lim_{h \to o} \frac{1}{h^2}(\phi(h) + \phi(-h) - 2\phi(0))$ et

$$\phi(h) + \phi(-h) - 2\phi(0) = \int (\mathrm{e}^{ihx} + \mathrm{e}^{-ihx} - 2)\, d\mu(x) = -4 \int \sin^2 \frac{hx}{2}\, d\mu(x).$$

Appliquant le lemme de Fatou (prop.1.3.4), on a

$$-\phi"(0) = \lim_h 4 \int \frac{\sin^2 \frac{hx}{2}}{h^2}\, d\mu(x) \geq 4 \int \underline{\lim}_h \frac{\sin^2 \frac{hx}{2}}{h^2 x^2} x^2\, d\mu(x) = \int x^2\, d\mu(x)$$ ∎

Fonctions caractéristiques usuelles.

a. Loi binomiale $B(n,p)$. Si $X \sim B(n,p)$, on a

$$\phi_X(t) = \mathbb{E}(\mathrm{e}^{itX}) = \sum_{k=0}^{n} C_n^k\, p^k (1-p)^{n-k} \mathrm{e}^{itk} = (p\mathrm{e}^{it} + 1 - p)^n.$$

Cette formule montre que, si $X \sim B(n,p)$ et $Y \sim B(m,p)$, X, Y indépendantes, alors $X + Y \sim B(n+m,p)$. En particulier si $X_1, \ldots, X_n$ sont des v.a. indépendantes avec $\mathbb{P}(X_k = 1) = p$, $\mathbb{P}(X_k = 0) = 1 - p$, $S_n = X_1 + \ldots + X_n \sim B(n,p)$.

b. Loi de Poisson $\mathcal{P}(\lambda)$. Si $X \sim \mathcal{P}(\lambda)$

$$\phi_X(t) = \mathbb{E}(\mathrm{e}^{itX}) = \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \mathrm{e}^{itk} = \exp(\lambda(\mathrm{e}^{it} - 1)).$$

Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$, X, Y indépendantes, alors $X + Y \sim \mathcal{P}(\lambda + \mu)$.

c. Loi gamma $G(a,c)$. Si $X \sim G(a,c)$, on a, $\phi_X(t) = \frac{c^a}{\Gamma(a)} \int_0^{+\infty} \mathrm{e}^{itx} \mathrm{e}^{-cx} x^{a-1}\, dx$. Utilisant la prop.1.3.7 et intégrant par partie, on obtient

$$\phi'_X(t) = \frac{ic^a}{\Gamma(a)} \int_0^{+\infty} \mathrm{e}^{itx} \mathrm{e}^{-cx} x^a\, dx = \frac{-iac^a}{\Gamma(a)(it - c)} \int_0^{+\infty} \mathrm{e}^{itx} \mathrm{e}^{-cx} x^{a-1}\, dx = \frac{ia}{c - it} \phi_X(t)$$

d'où $\phi_X(t) = (1 - \frac{it}{c})^{-a}$ puisque $\phi_X(0) = 1$. Noter que pour $a \notin \mathbb{N}$, on prend la détermination continue valant 1 en 0. Si $X \sim G(a,c)$ et $Y \sim G(b,c)$, X, Y indépendantes,

alors $X+Y \sim G(a+b,c)$. En particulier si $X_1,\ldots,X_n$ sont des v.a. indépendantes de même densité $\mathrm{e}^{-\lambda x}\mathbb{1}_{\{\mathbb{R}^+\}}$ et donc de loi $G(1,\lambda)$, $S_n = X_1+\ldots+X_n \sim G(n,\lambda)$ et a pour densité $\frac{\lambda^n}{(n-1)!}\mathrm{e}^{-\lambda x}x^{n-1}\mathbb{1}_{\{\mathbb{R}^+\}}$.

d. Loi normale $N_1(m,\sigma^2)$. On sait (lem.1.6.4) que, si $Y \sim N_1(0,1)$, $\phi_Y(t) = \mathrm{e}^{-\frac{t^2}{2}}$. Soit $X = m+\sigma Y$, alors $X \sim N_1(m,\sigma^2)$ et

$$\phi_X(t) = \mathbb{E}(\mathrm{e}^{itX}) = \mathbb{E}(\mathrm{e}^{it(m+\sigma Y)}) = \mathrm{e}^{itm}\mathbb{E}(\mathrm{e}^{it\sigma Y}) = \exp(itm - \frac{\sigma^2t^2}{2}).$$

On en déduit immédiatement

Proposition 2.7.8. *Si $X \sim N_1(m,\sigma^2)$ et $Y \sim N_1(l,\rho^2)$, X,Y indépendantes, alors $X+Y \sim N_1(m+l,\sigma^2+\rho^2)$.*

Transformation de Laplace.

Pour des v.a. positives, on peut avoir intérêt à utiliser la transformation de Laplace. Soit X une v.a. à valeurs $\mathbb{R}^+$ de loi μ_X. On appelle transformée de Laplace de X (ou plutôt de μ_X) la fonction sur $\mathbb{R}^+$

$$\psi_X(\lambda) = \mathbb{E}(\mathrm{e}^{-\lambda X}) = \int \mathrm{e}^{-\lambda x}\,d\mu_X(x). \tag{2.10}$$

L'espace vectoriel V des fonctions sur $\mathbb{R}^+$ de la forme $\sum c_i\mathrm{e}^{-\lambda_i x}$ est une algèbre séparant les points de $[0,+\infty]$, il résulte du théorème de Stone-Weierstrass que V est dense dans $C_0([0,+\infty[)$. On en déduit que $\psi_X = \psi_Y \Rightarrow \mu_X = \mu_Y$.
Appliquant la prop.1.3.7, on a, pour tout $\lambda > 0$, $\psi'_X(\lambda) = -\mathbb{E}(X\mathrm{e}^{-\lambda X})$ et, plus généralement, $\psi_X^{(n)}(\lambda) = (-1)^n\mathbb{E}(X^n\mathrm{e}^{-\lambda X})$. On en déduit que, si $\mathbb{E}(X) < +\infty$, $\mathbb{E}(X) = -\psi'_X(0)$, si $\mathbb{E}(X^2) < +\infty$, $\mathbb{E}(X^2) = \psi''_X(0)$,...
Enfin on vérifie sans peine que, si X et Y sont des v.a. positives indépendantes, $\psi_{X+Y}(\lambda) = \psi_X(\lambda)\psi_Y(\lambda)$.

Fonctions génératrices

Soit X une v.a. à valeurs $\mathbb{N}$ de loi $\mu_X(n) = \mathbb{P}(X=n)$. On appelle fonction génératrice de X la fonction définie pour $u \in [-1,+1]$ par

$$g_X(u) = \mathbb{E}(u^X) = \sum_{n=0}^{\infty}\mu_X(n)u^n, \tag{2.11}$$

où l'on a convenu que $u^0 \equiv 1$ (y compris pour $u=0$). Notons que $\psi_X(\lambda) = g_X(\mathrm{e}^{-\lambda})$, $\lambda \geq 0$.
Il résulte de la prop.1.3.7 (ou des résultats classiques sur les séries entières) que, pour $|u| < 1$, $g'_X(u) = \mathbb{E}(Xu^{X-1})$, $g''_X(u) = \mathbb{E}(X(X-1)u^{X-2})$, d'où l'on déduit que, si $\mathbb{E}(X) < +\infty$, $\mathbb{E}(X) = g'_X(1)$, si $\mathbb{E}(X^2) < +\infty$, $\mathbb{E}(X(X-1)) = g''_X(1)$,...
Ici encore on vérifie immédiatement que $g_{X+Y} = g_Xg_Y$. si X et Y sont des v.a. entières indépendantes.

On a également un critère d'indépendance. Si X et Y sont des v.a. entières, on appelle fonction génératrice de (X,Y), la fonction définie pour $u,v \in [-1,+1]$, par

$$g_{(X,Y)}(u,v) = \mathbb{E}(u^X v^Y) = \sum_{n=0}^{\infty} \mathbb{P}(X=n, Y=m) u^n v^m. \tag{2.12}$$

Alors les v.a. X et Y sont indépendantes ssi, pour tous u,v voisins de $(0,0)$, $g_{(X,Y)}(u,v) = g_X(u)g_Y(v)$.

2.7.2. Vecteurs gaussiens

On dit qu'une probabilité μ sur $\mathbb{R}$ est gaussienne si elle a pour densité (2.7) ou si $\mu = \delta_m$. Il est normal d'adjoindre les mesures de Dirac aux lois gaussiennes car (lem.1.6.5) la mesure $N_1(m,\sigma^2)$ converge étroitement vers δ_m lorsque $\sigma \to 0$. Une v.a. réelle est dite gaussienne si sa loi est gaussienne.

Définition 2.7.9. *Un vecteur aléatoire $X = (X_1, \ldots, X_d)$ est dit gaussien si, pour tout $a \in \mathbb{R}^d$, $a'X = a_1X_1 + \ldots + a_dX_d$ est une v.a.r. gaussienne.*

En particulier chaque composante X_k est une v.a.r. gaussienne mais cela ne suffit pas à assurer que le vecteur X est gaussien. On appelle loi gaussienne sur $\mathbb{R}^d$ toute loi d'un vecteur gaussien.

Exemples. (i) $X = 0 \in \mathbb{R}^d$ est un vecteur gaussien.

(ii) Soit $X = (X_1, \ldots, X_d)$ avec $X_1, \ldots, X_d$ indépendants de même loi $N_1(0,1)$. Alors (prop.2.7.8) $a_1X_1 + \ldots + a_dX_d \sim N_1(0, a_1^2 + \ldots + a_d^2)$ et X est un vecteur gaussien.

Cette notion est invariante par transformation linéaire, plus précisément:

Lemme 2.7.10. *Soit X un vecteur gaussien à valeurs $\mathbb{R}^d$ de moyenne m et de matrice de covariance K. Pour tous $b \in \mathbb{R}^r$ et M matrice $r \times d$, $Y = b + MX$ est un vecteur gaussien à valeurs $\mathbb{R}^r$ de moyenne $b + Mm$ et de matrice de covariance MKM'*

`Preuve:`
En effet $a'Y = a'b + (a'M)X$ est une v.a.r. gaussienne. On a $\mathbb{E}(Y) = b + M\mathbb{E}(X) = b + Mm$ et (prop.2.4.4) $K(Y) = K(MX) = MK(X)M' = MKM'$ ∎

Théorème 2.7.11. *Soit X un vecteur aléatoire de moyenne m et de matrice de covariance K. Le vecteur X est gaussien ssi sa fonction caractéristique est donnée par*

$$\phi_X(t) = \exp(it'm - \frac{1}{2}t'Kt). \tag{2.13}$$

`Preuve:`
(i) Supposons X gaussien. Alors (lem.2.7.10) $t'X \sim N_1(t'm, t'Kt)$ et $\phi_{t'X}(1) = \mathbb{E}(e^{it'X}) = \exp(it'm - \frac{1}{2}t'Kt)$ d'où (2.13).

(ii) Supposons (2.13). Alors $\phi_{a'X}(u) = \mathbb{E}(e^{iua'X}) = \exp(iua'm - \frac{1}{2}u^2a'Ka)$ donc $a'X$ est une v.a.r. gaussienne et X un vecteur gaussien ∎

Toute loi gaussienne sur $\mathbb{R}^d$ est donc déterminée par sa moyenne m et sa matrice de covariance K. On note $N_d(m, K)$ une telle loi. On a vu (exemple (ii)) que $N_d(0, I_d)$ existe mais on n'a pas établi l'existence dans le cas général. On a,

Lemme 2.7.12. *Soit K une matrice $d \times d$ symétrique semi-définie positive. Il existe une matrice $d \times d$ symétrique semi-définie positive A telle que $K = AA'$.*

`Preuve:`
Soient $\lambda_1, \dots, \lambda_d$ les valeurs propres de K qui sont ≥ 0. Il existe une matrice orthogonale C (i.e. $CC' = I$) telle que $C'KC = D = \text{diag}(\lambda_1, \dots, \lambda_d)$ où $\text{diag}(\lambda_1, \dots, \lambda_d)$ désigne la matrice diagonale ayant $\lambda_1, \dots, \lambda_d$ sur la diagonale. On a alors $CDC' = K$. Soit $\Delta = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$. On pose $A = C\Delta C'$. On a,

$$AA' = C\Delta C'(C\Delta C')' = C\Delta CC'\Delta C' = CDC' = K \quad \blacksquare$$

Appliquant le lem.2.7.10, on a que, si $X \sim N_d(0, I_d)$, $Y = m + AX \sim N_d(m, K)$. On a montré,

Théorème 2.7.13. *Etant donnés $m \in \mathbb{R}^d$ et une matrice $d \times d$ symétrique semi-définie positive K, il existe une et une seule loi gaussienne sur $\mathbb{R}^d$ de moyenne m et de matrice de covariance K.*

Vecteurs gaussiens et indépendance.

Théorème 2.7.14. *Soient $X = (X_1, \dots, X_d)$ un vecteur gaussien.*

(i) *Les v.a.r. $X_1, \dots, X_d$ sont indépendantes ssi la matrice de covariance $K(X)$ est diagonale.*

(ii) *On pose*

$$Y_1 = (X_1, \dots, X_{d_1}),\ Y_2 = (X_{d_1+1}, \dots, X_{d_2}), \dots Y_r = (X_{d_{r-1}+1}, \dots, X_d)'$$

Les vecteurs $(Y_1, \dots, Y_r)$ sont indépendants ssi $K_{i\,j}(X) = Cov(X_i, X_j) = 0$ pour tous i, j n'appartenant pas au même intervalle $[1, d_1]$, $[d_1+1, d_2], \dots, [d_{r-1}+1, d]$.

`Preuve:`
Seule la suffisance demande une preuve.

(i) Supposons $K(X)$ diagonale. On a $K(X) = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$ où $\sigma_k^2 = \text{Var}(X_k)$. Alors, notant $m = \mathbb{E}(X)$,

$$\phi_X(t) = \exp(i \sum_{k=1}^{d} m_k t_k - \frac{1}{2} \sum_{k=1}^{d} \sigma_k^2 t_k^2) = \prod_{k=1}^{d} \exp(i m_k t_k - \frac{1}{2} \sigma_k^2 t_k^2) = \phi_{X_1}(t_1) \dots \phi_{X_d}(t_d)$$

et donc (prop.2.7.5) les X_k sont indépendantes.

(ii) Supposons la condition sur les covariances réalisées. Elle implique, pour tous $u_1 \in \mathbb{R}^{d_1}, u_2 \in \mathbb{R}^{d_2-d_1}, \ldots$ et $p \neq q$, $\mathrm{Cov}(u'_pY_p, u'_qY_q) = 0$. Donc, d'après (i), les v.a.r. $u'_1Y_1, \ldots, u'_rY_r$ sont indépendantes. On a alors

$$\mathbb{E}(\mathrm{e}^{i(u'_1Y_1+\ldots+u'_rY_r)}) = \mathbb{E}(\mathrm{e}^{iu'_1Y_1})\ldots\mathbb{E}(\mathrm{e}^{iu'_rY_r})$$

et (prop.2.7.5) les v.a. $(Y_1, \ldots, Y_r)$ sont indépendantes ∎

Remarque. Attention à l'utilisation du th.2.7.14. On peut avoir X et Y v.a.r. gaussiennes, $\mathrm{Cov}(X,Y) = 0$ sans que les v.a. X et Y soient indépendantes. Par exemple si $X \sim N_1(0,1)$, si U est une v.a. indépendante de X telle que $\mathbb{P}(U = 1) = \mathbb{P}(U = -1) = \frac{1}{2}$ et si $Y = UX$, on vérifie facilement que $Y \sim N_1(0,1)$. On a $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) = \mathbb{E}(UX^2) = \mathbb{E}(U)\mathbb{E}(X^2) = 0$ et $|X| = |Y|$ donc X et Y ne sont pas indépendantes. En fait le couple (X,Y) n'est pas gaussien.

Le cas non dégénéré.

On dit que la loi $N_d(m,K)$ est non dégénéré si $\det(K) \neq 0$. Dans ce cas:

Théorème 2.7.15. *Si $X \sim N_d(m,K)$ et si $det(K) \neq 0$, X admet la densité*

$$h_{m,K}(x) = (2\pi)^{-\frac{d}{2}}(det(K))^{-\frac{1}{2}}\exp(-\frac{1}{2}(x-m)'K^{-1}(x-m)).$$

`Preuve:`
Soit A tel que $K = AA'$, on a $\det(A) = (\det(K))^{1/2}$ et A est inversible. Soit $Y \sim N_d(0, I_d)$ un vecteur gaussien de densité $(2\pi)^{-d/2}\exp(-\frac{|y|^2}{2})$. On a (lem.2.7.10) $Y = m + AY \sim N_d(m,K)$ et, pour $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\mathbb{E}(f(X)) = \mathbb{E}(f(m+AY)) = (2\pi)^{-\frac{d}{2}}\int f(m+Ay)\exp(-\frac{|y|^2}{2})\,dy.$$

On effectue le changement de variable $y = A^{-1}(x-m)$, on a $\frac{D(y)}{D(x)} = \det(A^{-1})$ et

$$\mathbb{E}(f(X)) = (2\pi)^{-\frac{d}{2}}\det(A^{-1})\int f(x)\exp(-\frac{1}{2}(x-m)'(A^{-1})'A^{-1}(x-m))\,dx.$$

Comme $K^{-1} = (AA')^{-1} = (A^{-1})'A^{-1}$, on a la formule annoncée ∎

2.7.3. Convergence en loi

On considère dans cette sous-section des v.a. à valeurs $\mathbb{R}^d$. Rappelons que si X est une telle v.a., on note μ_X sa loi et ϕ_X sa fonction caractéristique. Nous ferons un usage intensif des résultats de la section 1.7.

Définition 2.7.16. *On dit qu'une suite de v.a. X_n converge en loi vers une probabilité μ (resp. une v.a. X) si la suite μ_{X_n} converge étroitement vers μ (resp. vers μ_X).*

La distinction entre convergence en loi vers μ ou vers X est une simple affaire de langage car en fait c'est la loi de X_n qui converge vers μ et donc vers la loi de X pour toute v.a. X de loi μ. Ceci signifie donc que, pour toute $f \in C_b$, $\mathbb{E}(f(X_n)) = \int f\, d\mu_{X_n} \to_n \int f\, d\mu$ $(= \mathbb{E}(f(X))$. D'après la prop.1.7.2 il suffit de le vérifier pour $f \in C_k$ et d'après les th.1.7.9 et 1.7.10, on a:

Théorème 2.7.17. *La suite X_n converge en loi vers μ (resp. vers X) ssi, pour tout $t \in \mathbb{R}^d$, $\phi_{X_n}(t) \to_n \hat{\mu}(t)$ (resp. $\phi_X(t)$).*

Théorème 2.7.18. *Soit X_n une suite de v.a. telle que, pour toute $t \in \mathbb{R}^d$, $\phi_{X_n}(t) \to_n \psi(t)$. Si ψ est continue en 0, il existe une probabilité μ sur $\mathbb{R}^d$ telle que $\psi = \hat{\mu}$ et X_n converge en loi vers μ.*

Examinons le lien entre la convergence en loi et les convergences des v.a. étudiées dans la section précédente.

Z. La grande différence entre cette notion de convergence et les précédentes (convergence en probabilité , convergence p.s., convergence dans L^p) est que:

- Cette notion de convergence n'est pas de type “produit”, c'est à dire que si X_n et Y_n convergent en loi vers X et Y, il se peut très bien que le couple (X_n, Y_n) ne converge pas en loi...

- Cette notion de convergence n'est pas de type “espace vectoriel”, c'est à dire que si X_n et Y_n convergent en loi vers X et Y, on ne peut, en général, rien dire sur la convergence en loi d'une combinaison linéaire de X_n et Y_n.

D'ailleurs, la convergence en loi n'est pas à proprement parler, une notion de convergence de variables aléatoires , mais, comme son nom l'indique, une notion de convergence de suites de lois de probabilité .

Proposition 2.7.19. *Si X_n converge en probabilité vers X, alors X_n converge en loi vers X.*

`Preuve:`
Il suffit (prop.1.7.2) de montrer que, pour toute $f \in C_k$, $\mathbb{E}(f(X_n)) \to_n \mathbb{E}(f(X))$. Soient donc $f \in C_k$ et $\varepsilon > 0$. Il existe, f étant uniformément continue, $\alpha > 0$ tel que $|f(x) - f(y)| \le \varepsilon$ si $|x - y| \le \alpha$. On a alors

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \le \mathbb{E}(|f(X_n)) - f(X)|\mathbb{1}_{\{|X_n - X| \le \alpha\}})$$
$$+\mathbb{E}(|f(X_n)) - f(X)|\mathbb{1}_{\{|X_n - X| > \alpha\}}) \le \varepsilon + 2||f||\mathbb{P}(|X_n - X| > \alpha)$$

d'où $\overline{\lim}_n |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \le \varepsilon$ et $\mathbb{E}(f(X_n)) \to_n \mathbb{E}(f(X))$ ∎

Exemple. Soir X_n une suite de v.a.r. telle que $\mathbb{P}(X_n = 1) = p_n$ et $\mathbb{P}(X_n = 0) = 1 - p_n$ avec $0 < p_n < 1$. $X_n \to_n 0$ en probabilité ssi $p_n \to_n 0$, $X_n \to_n 1$ en probabilité ssi $p_n \to_n 1$ et sinon ne converge pas en probabilité tandis que, vu que $\mathbb{E}(f(X_n)) =$

$p_n f(1) + (1 - p_n) f(0)$, X_n converge en loi ssi $p_n \to_n p$. Ceci montre qu'en général la convergence en loi n'implique pas la convergence en probabilité.

A titre d'application, on démontre un résultat utile sur les v.a.r. gaussiennes.

Lemme 2.7.20. *Soit a_n une suite de réels tels que, pour tout $t \in \mathbb{R}$, $e^{ita_n} \to_n \lambda(t)$. Alors la suite a_n converge.*

`Preuve:`
La fonction $\lambda(t)$ est continue et $|\lambda(t)| = 1$, il existe donc $\alpha > 0$ avec $\int_0^\alpha \lambda(t)\, dt = \rho \neq 0$. D'après le théorème de Lebesgue, $\lim_n \int_0^\alpha \mathrm{e}^{ita_n}\, dt = \rho$ et, comme $\int_0^\alpha ia_n \mathrm{e}^{ita_n}\, dt = \mathrm{e}^{i\alpha a_n} - 1$, pour n assez grand:

$$a_n = \frac{\mathrm{e}^{i\alpha a_n} - 1}{i \displaystyle\int_0^\alpha \mathrm{e}^{ita_n}\, dt} \to a = \frac{\lambda(\alpha)}{i\rho} \quad \blacksquare$$

Proposition 2.7.21. *Soit X_n une suite de v.a.r. gaussiennes convergeant en probabilité vers X. Alors X est gaussienne et, pour tout $p \geq 1$, X_n converge vers X dans L^p.*

`Preuve:`
Soient $a_n = E(X_n)$, $\sigma_n^2 = \mathrm{Var}(X_n)$. On a $\phi_{X_n}(t) = \exp(ia_n t - \sigma_n^2 t^2/2)$. Comme X_n converge vers X en probabilité, X_n converge vers X en loi et $\phi_{X_n}(t)$ converge vers la f.c. $\phi(t)$ de X. Alors $|\phi_{X_n}(t)| = \exp(-\sigma_n^2 t^2/2) \to |\phi(t)|$. On a $\phi(1) \neq 0$, sinon $\sigma_n^2 \to +\infty$ et, pour tout $t \neq 0$, $\phi(t) = 0$, ce qui est impossible puisque ϕ est une fonction caractéristique. Donc $\sigma_n^2 \to -2 \ln |\phi(1)| = \sigma^2 < +\infty$. Alors, pour tout $t \in \mathbb{R}$, $\exp(ia_n t)$ converge et (lem.2.7.20) $a_n \to a$. On a donc $\phi(t) = \exp(iat - \sigma^2 t^2/2)$ et $X \sim N_1(a, \sigma^2)$. Vu que

$$\mathbb{E}(e^{|X_n|}) \leq E(e^{X_n}) + E(e^{-X_n}) = \exp(a_n + \sigma_n^2/2) + \exp(-a_n + \sigma_n^2/2)$$

et que les suites a_n et σ_n^2 sont bornées, on a $\sup_n E(e^{|X_n|}) < +\infty$. Comme, pour tout $r \in \mathbb{N}$, $|x|^r \leq C_r e^{|x|}$, ceci implique que $\sup_n E|X_n|^r < +\infty$. Il suffit alors d'appliquer le thm.2.6.5 $\blacksquare$

La prop.1.7.5 et le th.1.7.8 deviennent:

Proposition 2.7.22. *Soit X_n une suite de v.a. convergeant en loi vers μ. Pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mu(\partial A) = 0$, on a $\mathbb{P}(X_n \in A) \to_n \mu(A)$.*

Théorème 2.7.23. *Soit X_n une suite de v.a. telle que, pour tout $\varepsilon > 0$, il existe un compact K tel que, pour tout n, $\mathbb{P}(X_n \in K) \geq 1 - \varepsilon$, alors il existe une sous-suite X_{n_k} convergeant en loi.*

Le théorème de la limite centrale.

Théorème 2.7.24. *Soit X_n une suite de v.a. de L_d^2 indépendantes et de même loi. On pose $m = \mathbb{E}(X_1)$, $K = K(X_1)$, $S_n = X_1 + \ldots, X_n$. Alors $\frac{1}{\sqrt{n}}(S_n - nm)$ converge en loi vers $N_d(0, K)$.*

`Preuve:`
On peut supposer $\mathbb{E}(X_1)=0$. Soit $Y\sim N_d(0,K)$. On a $\phi_{\frac{1}{\sqrt{n}}S_n}(t)=(\phi_{X_1}(\frac{t}{\sqrt{n}}))^n$. De plus, vu la prop.2.7.6, $\frac{\partial}{\partial t_k}\phi_{X_1}(0)=0=\frac{\partial}{\partial t_k}\phi_Y(0)$, $\frac{\partial^2}{\partial t_j t_k}\phi_{X_1}(0)=-K_{j,k}=\frac{\partial^2}{\partial t_j t_k}\phi_Y(0)$. On a donc

$$\phi_{X_1}(t)-\phi_Y(t)=|t|^2\varepsilon(t),\quad |\varepsilon(t)|\to_{t\to 0}0.$$

Si $u,v\in\mathbb{C}$ avec $|u|\le 1$, $|v|\le 1$, on a $|u^n-v^n|\le n|u-v|$ et donc, vu que $\phi_Y(t)=\exp(-\frac{1}{2}t'Kt)=(\phi_Y(\frac{t}{\sqrt{n}}))^n$,

$$|\phi_{\frac{1}{\sqrt{n}}S_n}(t)-\phi_Y(t)|=|(\phi_{X_1}(\frac{t}{\sqrt{n}}))^n-(\phi_Y(\frac{t}{\sqrt{n}}))^n|$$
$$\le n|\phi_{X_1}(\frac{t}{\sqrt{n}})-\phi_Y(\frac{t}{\sqrt{n}})|\le n\frac{|t|^2}{n}|\varepsilon(\frac{t}{\sqrt{n}})|\le|t|^2|\varepsilon(\frac{t}{\sqrt{n}})|\to_n 0$$

et $\frac{1}{\sqrt{n}}S_n$ converge en loi vers $N_d(0,K)$ d'après le th.2.7.17 ∎

Corollaire 2.7.25. *Soit X_n une suite de v.a.r. indépendantes, de même loi, de carré intégrable. On pose $m=\mathbb{E}(X_1)$, $\sigma^2=\mathrm{Var}(X_1)$, $S_n=X_1+\ldots,X_n$. Alors, pour $-\infty\le a<b\le+\infty$,*

$$\mathbb{P}(a<\frac{S_n-nm}{\sigma\sqrt{n}}<b)\to_n\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{t^2}{2}}\,dt.$$

`Preuve:`
Ceci résulte du th.2.7.24 et de la prop.2.7.22 ∎

Convergence en loi et fonction de répartition.

Soit F une fonction de répartition sur $\mathbb{R}$. Rappelons que F croît de 0 à 1 et que F est continue à droite. On note $C(F)$ l'ensemble des points de continuité de F. Noter que $C^c(F)$ est au plus dénombrable.

Théorème 2.7.26. *Soient μ_n et μ des probabilités sur $\mathbb{R}$ de fonction de répartition F_n et F. Alors μ_n converge étroitement vers μ ssi, pour tout $x\in C(F)$, $F_n(x)\to_n F(x)$.*

`Preuve:`
(i) On suppose que μ_n converge étroitement vers μ. Soit $x\in C(F)$. On a $\mu(\{x\})=F(x)-F(x-)=0$ et (prop.1.7.5) $F_n(x)=\mu_n(]-\infty,x])\to_n\mu(]-\infty,x])=F(x)$.

(ii) On suppose que, pour tout $x\in C(F)$, $F_n(x)\to_n F(x)$. Montrons d'abord que la suite μ_n est tendue (def.1.7.7). Soit $\varepsilon>0$. Il existe $a<b$ tels que $F(b)-F(a)>1-\varepsilon$ et, $C^c(F)$ étant au plus dénombrable, on peut supposer $a,b\in C(F)$. On a alors $F_n(b)-F_n(a)\to_n F(b)-F(a)$ et il existe n_0 tel que, pour tout $n\ge n_0$, $F_n(b)-F_n(a)>1-\varepsilon$ i.e. $\mu_n([a,b]^c)<\varepsilon$. La suite μ_n est donc tendue.
Supposons que μ_{n_k} converge étroitement vers ν. Vu (i), on a $F(x)=F_\nu(x)$ pour tout $x\in C(F)\cap C(F_\nu)$. Ce dernier ensemble étant dense (son complémentaire est au plus dénombrable), on a $F=F_\nu$ et $\mu=\nu$.
Ces deux propriétés impliquent la convergence étroite de μ_n vers μ. Sinon il existerait $f\in C_b$ telle que $\mu_n(f)$ ne converge pas vers $\mu(f)$ et donc $\alpha>0$ et n_k tels que, pour tout

k, $|\mu_{n_k}(f) - \mu(f)| > \alpha$. Mais la suite μ_{n_k} étant tendue, on peut en extraire (th.1.7.8) une sous-suite convergeant étroitement et nécessairement vers μ d'où une contradiction

∎

Chapitre 3

Espérances conditionnelles

3.1. Définition élémentaire

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Soit B un événement tel que $\mathbb{P}(B) > 0$. Pour $A \in \mathcal{A}$, on appelle probabilité conditionnelle de A sachant B la quantité $\mathbb{P}(A|B) := \mathbb{P}(A \cap B)/\mathbb{P}(B)$ et, pour X v.a. intégrable, espérance conditionnelle de X sachant B la quantité

$$\mathbb{E}(X|B) := \frac{1}{\mathbb{P}(B)} \int_B X \, d\mathbb{P}.$$

Il suffit de s'intéresser à la deuxième quantité puisque $\mathbb{P}(A|B) = \mathbb{E}(\mathbb{1}_{\{A\}}|B)$. Noter que $\mathbb{E}(X|B)$ représente la nouvelle espérance de X sachant que le point tiré $\omega \in B$. En fait on sait si le point tiré $\omega \in B$ ou $\omega \in B^c$, c'est à dire qu'on connaît la tribu $\mathcal{B} = \{\Omega, \emptyset, B, B^c\}$. On appellera donc espérance conditionnelle de X sachant $\mathcal{B}$ et on notera $\mathbb{E}(X|\mathcal{B})$ la v.a.

$$Y = \left(\frac{1}{\mathbb{P}(B)} \int_B X \, d\mathbb{P}\right) \mathbb{1}_{\{B\}} + \left(\frac{1}{\mathbb{P}(B^c)} \int_{B^c} X \, d\mathbb{P}\right) \mathbb{1}_{\{B^c\}}.$$

Supposons maintenant que $\mathcal{B} = \sigma(B_k, \ k \geq 0)$ où les $B_k \in \mathcal{A}$ forment une partition de Ω avec $\mathbb{P}(B_k) > 0$. Rappelons qu'alors une v.a. Z est $\mathcal{B}$-mesurable ssi $Z = \sum_k c_k \mathbb{1}_{\{B_k\}}$. Dans ce cas, X étant une v.a. intégrable, on appelle espérance conditionnelle de X sachant $\mathcal{B}$ et on note $\mathbb{E}(X|\mathcal{B})$ ou encore $\mathbb{E}^{\mathcal{B}}(X)$, la v.a.

$$Y = \sum_k \left(\frac{1}{\mathbb{P}(B_k)} \int_{B_k} X \, d\mathbb{P}\right) \mathbb{1}_{\{B_k\}}. \tag{3.1}$$

On a dans ce cadre la caractérisation suivante

Proposition 3.1.1. *La v.a. Y définie par (3.1) est l'unique v.a. $\mathcal{B}$-mesurable telle que*

$$\text{pour tout } B \in \mathcal{B}, \quad \int_B Y \, d\mathbb{P} = \int_B X \, d\mathbb{P}. \tag{3.2}$$

`Preuve:`
Puisque tout $B \in \mathcal{B}$ est réunion disjointe de B_k, pour montrer que Y vérifie (3.2), il suffit de le vérifier pour $B = B_k$ ce qui est immédiat. Réciproquement soit Z une v.a. $\mathcal{B}$-mesurable vérifiant (3.2). On a $Z = \sum_k c_k \mathbb{1}_{\{B_k\}}$ et $\int_{B_k} Z\, d\mathbb{P} = c_k \mathbb{P}(B_k) = \int_{B_k} X\, d\mathbb{P}$ d'où $c_k = \frac{1}{\mathbb{P}(B_k)} \int_{B_k} X\, d\mathbb{P}$ et $Z = Y$ p.s. ∎

On arrive au problème suivant. Soient $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ et X une v.a. intégrable; existe-t-il une v.a. Y, $\mathcal{B}$-mesurable, vérifiant (3.2) et est-elle unique?

3.2. Définition et propriétés

Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$.

Dans la suite on utilisera souvent le résultat suivant.

Proposition 3.2.1. *Si X et Y sont deux v.a.r. $\mathcal{B}$-mesurables toutes deux positives ou toutes deux intégrables, vérifiant, pour tout $B \in \mathcal{B}$, $\int_B X\, dP \geq \int_B Y\, dP$ (resp. $=$), alors $X \geq Y$ p.s. (resp. $=$).*

`Preuve:`
Soit $F_{a,b} = \{X \leq a < b \leq Y\} \in \mathcal{B}$. Puisque $\{X < Y\} = \cup_{a,b \in \mathbf{Q}} F_{a,b}$, il suffit de montrer que, pour tous $a < b$, $\mathbb{P}(F_{a,b}) = 0$. Mais si $\mathbb{P}(F_{a,b}) > 0$,

$$\int_{F_{a,b}} X\, dP \leq a\mathbb{P}(F_{a,b}) < b\mathbb{P}(F_{a,b}) \leq \int_{F_{a,b}} Y\, dP,$$

ce qui contredit l'hypothèse ∎

On peut maintenant énoncer le résultat principal de ce chapitre. Commençons par le cas L^2.

Théorème 3.2.2. *Soit X une v.a.r. de carré intégrable. Il existe une v.a.r. Y de carré intégrable, $\mathcal{B}$-mesurable, unique à une équivalence près, telle que*

$$\textit{pour toute v.a. } Z \textit{ de carré intégrable } \mathcal{B}\textit{-mesurable,} \quad \mathbb{E}(YZ) = \mathbb{E}(XZ). \tag{3.3}$$

`Preuve:`
Remarquons d'abord que l'unicité résulte de la prop.3.2.1. Montrons l'existence. Soit $X \in L^2$ (avec l'abus de langage usuel). $H = L^2(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de Hilbert et $K = \{Z \in H,\ Z \text{ a un représentant } \mathcal{B}\text{-mesurable}\}$ est un sous espace fermé de H puisque $K \simeq L^2(\Omega, \mathcal{B}, \mathbb{P})$. Soit Y (un représentant de) la projection orthogonale de X sur K. On a, pour tout $Z \in K$, $X - Y \perp K$ i.e. $\mathbb{E}(YZ) = \mathbb{E}(XZ)$ ∎

Théorème 3.2.3. *Soit X une v.a.r. intégrable (resp. positive). Il existe une v.a.r. Y intégrable (resp. positive), $\mathcal{B}$-mesurable, unique à une équivalence près, telle que*

$$\textit{pour tout } B \in \mathcal{B}, \quad \int_B Y\, dP = \int_B X\, dP. \tag{3.4}$$

`Preuve:`
Ici encore l'unicité résulte de la prop.3.2.1. Montrons l'existence. On suppose $X \geq 0$. Soit $X_n = X \wedge n$, $X_n \in L^2$. Il existe donc Y_n $\mathcal{B}$-mesurable tel que, pour tout $B \in \mathcal{B}$, $\int_B Y_n\, dP = \int_B X_n\, dP$. La prop.3.2.1 implique que Y_n est positif et que $Y_n \leq Y_{n+1}$. On pose $Y = \lim \uparrow Y_n$, Y est $\mathcal{B}$-mesurable et (Beppo-Levi), pour tout $B \in \mathcal{B}$, $\int_B Y\, dP = \lim \uparrow \int_B Y_n\, dP = \lim \uparrow \int_B X_n\, dP = \int_B X\, dP$. Notons que si X est intégrable, Y également et donc $Y < +\infty$ p.s. Si $X \in L^1$, on écrit $X = X^+ - X^-$ ∎

Remarque 1. On peut donner une démonstration directe du th.3.2.3 en utilisant le théorème de Radon-Nikodym (th.1.8.7). Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Posons, pour $B \in \mathcal{B}$, $\nu_X(B) = \int_B X\, d\mathbb{P}$. ν_X est une mesure signée sur $(\Omega, \mathcal{B})$ et, notant $\mathbb{P}_\mathcal{B}$ la restriction de $\mathbb{P}$ à $\mathcal{B}$, on a, pour tout $B \in \mathcal{B}$, $\mathbb{P}_\mathcal{B}(B) = 0$ implique $\nu_X(B) = 0$. On peut appliquer le th.1.8.7 sur l'espace $(\Omega, \mathcal{B})$. Il existe donc $Y \in L^1(\Omega, \mathcal{B}, \mathbb{P})$ unique telle que, pour tout $B \in \mathcal{B}$,

$$\nu_X(B) = \int_B X\, d\mathbb{P} = \int_B Y\, d\mathbb{P}_\mathcal{B} = \int_B Y\, d\mathbb{P}.$$

On a donc

$$\mathbb{E}(X|\mathcal{B}) = \frac{d\,\nu_X}{d\,\mathbb{P}_\mathcal{B}}.$$

La v.a. Y définie par le th.3.2.3 s'appelle l'espérance conditionnelle de X sachant $\mathcal{B}$ et se note $\mathbb{E}(X|\mathcal{B})$ ou $\mathbb{E}^\mathcal{B}(X)$. Si $X = \mathbb{1}_{\{A\}}$, $A \in \mathcal{A}$, $\mathbb{E}(\mathbb{1}_{\{A\}}|\mathcal{B})$ se note $\mathbb{P}(A|\mathcal{B})$ et s'appelle la probabilité conditionnelle de A sachant $\mathcal{B}$. Si $X = (X_1, \ldots, X_d) \in L^1_d$, le vecteur aléatoire $(\mathbb{E}(X_1|\mathcal{B}), \ldots, \mathbb{E}(X_d|\mathcal{B}))$ s'appelle encore l'espérance conditionnelle de X sachant $\mathcal{B}$ et se note aussi $\mathbb{E}(X|\mathcal{B})$.

Remarque 2: Si on écrit $X = X^+ - X^-$, $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X^+|\mathcal{B}) - \mathbb{E}(X^-|\mathcal{B})$ est définie sans ambiguïté dès que $\mathbb{E}(X^+) < +\infty$ ou $\mathbb{E}(X^-) < +\infty$.

Proposition 3.2.4. *Du th.3.2.3 et de la prop.3.2.1 résulte immédiatement:*

- $\mathbb{E}(1|\mathcal{B}) = 1$ *p.s. et, si* $\mathcal{B} = \{\Omega, \emptyset\}$, $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$ *($X \geq 0$ ou intégrable).*
- *Pour $X, Y \geq 0$ et $a, b \geq 0$ (ou X, Y intégrables et $a, b \in \mathbb{R}$),* $\mathbb{E}(aX + bY|\mathcal{B}) = a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B})$ *p.s.*
- *Pour $X, Y \geq 0$ (ou X, Y intégrables), $X \leq Y$ p.s. implique* $\mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(Y|\mathcal{B})$ *p.s.*

On a aussi prenant $B = \Omega$ dans (3.4):

Proposition 3.2.5. *Pour X positive ou intégrable,* $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$.

On utilisera très souvent le résultat suivant, établi dans le premier chapitre. Une v.a. Z est dite $\mathcal{B}$-mesurable étagée si elle est de la forme $Z = \sum_{k=1}^n c_k \mathbb{1}_{\{B_k\}}$, $B_k \in \mathcal{B}$. On sait (prop.1.1.2) que toute v.a. $\mathcal{B}$-mesurable positive est limite d'une suite croissante de v.a. $\mathcal{B}$-mesurables étagées positives. Rappelons aussi qu'on note $\mathcal{B}^+$ l'ensemble des v.a. $\mathcal{B}$-mesurables positives, $\mathcal{B}_b$ l'ensemble des v.a.r. $\mathcal{B}$-mesurables bornées et que l'on désigne l'espérance conditionnelle par $\mathbb{E}^\mathcal{B}(X)$ ou bien par $\mathbb{E}(X|\mathcal{B})$.

Proposition 3.2.6. *Soit X une variable aléatoire positive (resp. variable aléatoire intégrable). Alors $\mathbb{E}^{\mathcal{B}}(X)$ vérifie*

$$\boxed{\mathbb{E}(Z\mathbb{E}^{\mathcal{B}}(X)) = \mathbb{E}(ZX)} \tag{3.5}$$

pour toute $Z \in \mathcal{B}^+$ (resp $Z \in \mathcal{B}_b$).

`Preuve:`
Supposons d'abord $X \geq 0$. La relation $\mathbb{E}(ZY) = \mathbb{E}(ZX)$ est, par définition, vraie pour $Z = \mathbb{1}_{\{B\}}$, $B \in \mathcal{B}$. Par linéarité (prop.3.2.5), elle est vraie pour Z étagée $\mathcal{B}$-mesurable positive puis (prop.1.1.2 et Beppo-Levi) pour Z $\mathcal{B}$-mesurable positive. Si $X \in L^1$, on se ramène au cas positif en écrivant $X = X^+ - X^-$ et $Z = Z^+ - Z^-$ ∎

Proposition 3.2.7. *Soit $(\mathcal{B}_i;\ i \in I)$ une famille finie ou dénombrable de sous tribus de $\mathcal{A}$ et $\mathcal{B} = \sigma(\mathcal{B}_i;\ i \in I)$. Soit X une variable aléatoire intégrable sur $(\Omega, \mathcal{A}, \mathbb{P})$. Pour qu'une variable aléatoire $\mathcal{B}$ mesurable et intégrable U soit une version de $\mathbb{E}^{\mathcal{B}}(X)$, il suffit que l'on ait:*

$$\mathbb{E}(U\prod_{i\in J} Z_i) = \mathbb{E}(X\prod_{i\in J} Z_i), \quad \textit{ou bien} \quad \int_{\cap_{i\in J}B_i} U\, d\mathbb{P} = \int_{\cap_{i\in J}B_i} X\, d\mathbb{P}$$

pour tout sous ensemble fini $J \subset I$ et tous $Z_i \in (\mathcal{B}_i)_b$ ou bien $B_i \in \mathcal{B}_i$.

`Preuve:`
Il suffit de considérer le second cas. On définit $\mathcal{C} = \{\cap_{i\in J}B_i\ ;\ B_i \in \mathcal{B}_i\ ,\ J \text{ fini}\}$ et $\mathcal{M} = \{A \in \mathcal{A}\ ;\ \mathbb{E}(U\mathbb{1}_{\{A\}}) = \mathbb{E}(X\mathbb{1}_{\{A\}})\}$. On vérifie facilement les hypothèses de la prop.1.1.1. On a donc $\sigma(\mathcal{C}) = \mathcal{B} \subset \mathcal{M}$ d'où, pour tout $B \in \mathcal{B}$, $\mathbb{E}(U\mathbb{1}_{\{B\}}) = \mathbb{E}(X\mathbb{1}_{\{B\}})$ i.e. $\mathbb{E}(X|\mathcal{B}) = U$ p.s. ∎

Lorsque la tribu $\mathcal{B}$ est engendrée par un système fini ou dénombrable de variables aléatoires , on obtient une caractérisation de l'espérance conditionnelle qui sera d'usage constant dans le chapitre consacré aux chaînes de Markov.

Corollaire 3.2.8. *Soit $(Y_i;\ i \in I)$ une famille finie ou dénombrable de variables aléatoires sur $(\Omega, \mathcal{A})$ et à valeurs dans des espaces mesurables $(E_i, \mathcal{E}_i)$, $\mathcal{B} = \sigma(Y_i;\ i \in I)$ la tribu engendrée par cette famille. Soit X une variable aléatoire intégrable sur $(\Omega, \mathcal{A}, \mathbb{P})$. Pour qu'une variable aléatoire $\mathcal{B}$ mesurable et intégrable U soit une version de $\mathbb{E}^{\mathcal{B}}(X)$, il suffit que l'on ait:*

$$\mathbb{E}(U\prod_{i\in J} h_i(Y_i)) = \mathbb{E}(X\prod_{i\in J} h_i(Y_i)), \quad \textit{ou bien} \quad \int_{\cap_{i\in J}(Y_i\in B_i)} U\, d\mathbb{P} = \int_{\cap_{i\in J}(Y_i\in B_i)} X\, d\mathbb{P}$$

pour tout sous ensemble fini $J \subset I$ et tous $h_i \in (\mathcal{E}_i)_b$ ou bien $B_i \in \mathcal{E}_i$.

Etablissons deux propriétés plus spécifiques des espérances conditionnelles.

Proposition 3.2.9. *Soient U et V des v.a. positives (ou telles que UV et V soient intégrables) avec U $\mathcal{B}$-mesurable, alors $\mathbb{E}^{\mathcal{B}}(UV) = U\mathbb{E}^{\mathcal{B}}(V)$ p.s.*

`Preuve:`
Considérons le cas positif. Soient $Y = U\mathbb{E}^{\mathcal{B}}(V)$ et $Z \in \mathcal{B}^+$. Alors $Y \in \mathcal{B}^+$ et on a (prop.3.2.6) $\mathbb{E}(ZY) = \mathbb{E}(ZU\mathbb{E}^{\mathcal{B}}(V)) = \mathbb{E}(ZUV)$ i.e. $Y = \mathbb{E}^{\mathcal{B}}(UV)$. p.s. Le cas intégrable se traite en écrivant $U = U^+ - U^-$ et $V = V^+ - V^-$ ■

Proposition 3.2.10. *Soient $\mathcal{C} \subset \mathcal{B}$ des sous-tribus de $\mathcal{A}$ et X une v.a. positive ou intégrable, alors $E^{\mathcal{C}}(X) = E^{\mathcal{C}}(\mathbb{E}^{\mathcal{B}}(X))$. p.s.*

`Preuve:`
Il suffit de considérer le cas positif. Posons $Y = E^{\mathcal{C}}(\mathbb{E}^{\mathcal{B}}(X))$, $Y \in \mathcal{C}^+$. Soit $Z \in \mathcal{C}^+ \subset \mathcal{B}^+$. On a

$$\mathbb{E}(ZY) = \mathbb{E}(ZE^{\mathcal{C}}(\mathbb{E}^{\mathcal{B}}(X))) = \mathbb{E}(Z\mathbb{E}^{\mathcal{B}}(X)) = \mathbb{E}(ZX)$$

i.e. $Y = E^{\mathcal{C}}(X)$ ■

Généralisation des inégalités de Schwarz et de Jensen aux espérances conditionnelles.

Proposition 3.2.11. *Soient $X, Y \in L^2$. Alors $|\mathbb{E}^{\mathcal{B}}(XY)|^2 \leq \mathbb{E}^{\mathcal{B}}(X^2)\mathbb{E}^{\mathcal{B}}(Y^2)$ p.s.*

`Preuve:`
On a, sauf sur un ensemble négligeable N, pour tout $\lambda \in \mathbb{Q}$, $0 \leq \mathbb{E}^{\mathcal{B}}(X + \lambda Y)^2 = \mathbb{E}^{\mathcal{B}}(X^2) - 2\lambda\mathbb{E}^{\mathcal{B}}(XY) + \lambda^2\mathbb{E}^{\mathcal{B}}(Y^2)$ d'où l'inégalité en écrivant que $\Delta \leq 0$ p.s. ■

Proposition 3.2.12. *Soit $f : \mathbb{R} \to \mathbb{R}$ convexe. On suppose X et $f(X)$ intégrables. Alors $f(\mathbb{E}^{\mathcal{B}}(X)) \leq \mathbb{E}^{\mathcal{B}}(f(X))$ p.s.*

`Preuve:`
Soit x_n une suite dense dans $\mathbb{R}$. Reprenant l'argument de la prop.2.3.4, pour chaque n, il existe une fonction affine $\alpha_n(x) = a_n x + b_n \leq f(x)$ telle que $\alpha_n(x_n) = f(x_n)$. Soit $g = \sup_n \alpha_n$, g est convexe (tout sup de fonctions convexes est convexe), continue (toute fonction convexe sur $\mathbb{R}$ est continue) et, pour tout n, $g(x_n) = f(x_n)$ donc $g = f$ i.e. $f(x) = \sup_n(a_n x + b_n)$. On a alors, pour tout n, $a_n X + b_n \leq f(X)$ d'où p.s. $a_n\mathbb{E}^{\mathcal{B}}(X) + b_n \leq \mathbb{E}^{\mathcal{B}}(f(X))$ et $f(\mathbb{E}^{\mathcal{B}}(X)) = \sup_n(a_n\mathbb{E}^{\mathcal{B}}(X) + b_n) \leq \mathbb{E}^{\mathcal{B}}(f(X))$ ■

Soit $p \geq 1$. La fonction $x \mapsto |x|^p$ étant convexe, on a, pour $X \in L^p$, $|\mathbb{E}^{\mathcal{B}}(X)|^p \leq \mathbb{E}^{\mathcal{B}}(|X|^p)$ et prenant l'espérance, $\mathbb{E}(|\mathbb{E}^{\mathcal{B}}(X)|^p) \leq \mathbb{E}(\mathbb{E}^{\mathcal{B}}(|X|^p)) = \mathbb{E}(|X|^p)$ d'où

$$\text{pour } p \geq 1, \quad ||\mathbb{E}^{\mathcal{B}}(X)||_p \leq ||X||_p \tag{3.6}$$

ce qui signifie que l'opérateur $X \mapsto \mathbb{E}^{\mathcal{B}}(X)$ est une contraction de L^p.

Il reste à étudier les passages à la limite. En fait, on a les mêmes résultats que pour l'espérance ordinaire.

Proposition 3.2.13. *Soit X_n une suite de variables aléatoires .*

(i) *Si $0 \leq X_n \uparrow X$, $\mathbb{E}^{\mathcal{B}}(X_n) \uparrow \mathbb{E}^{\mathcal{B}}(X)$ p.s.*

(ii) *Si* $0 \leq X_n$, $\mathbb{E}^{\mathcal{B}}(\underline{\lim}\, X_n) \leq \underline{\lim}\, \mathbb{E}^{\mathcal{B}}(X_n)$ *p.s.*

(iii) *Si* $|X_n| \leq V$ *avec* $V \in L^1$,

$$\mathbb{E}^{\mathcal{B}}(\underline{\lim}\, X_n) \leq \underline{\lim}\, \mathbb{E}^{\mathcal{B}}(X_n) \leq \overline{\lim}\, \mathbb{E}^{\mathcal{B}}(X_n) \leq \mathbb{E}^{\mathcal{B}}(\overline{\lim}\, X_n) \text{ p.s.}$$

(iv) *Si* $|X_n| \leq V$ *avec* $V \in L^1$ *et si* $X_n \to X$ *p.s.*, $\mathbb{E}^{\mathcal{B}}(X_n) \to \mathbb{E}^{\mathcal{B}}(X)$ *p.s.*

`Preuve:`
(i) Vu la prop.3.2.4, $\mathbb{E}^{\mathcal{B}}(X_n) \uparrow$ p.s. On pose $Y = \overline{\lim}\, \mathbb{E}^{\mathcal{B}}(X_n)$, Y est $\mathcal{B}$-mesurable et on a $\mathbb{E}^{\mathcal{B}}(X_n) \uparrow Y$ p.s. Vu que pour tout $B \in \mathcal{B}$, $\int_B X_n\, d\mathbb{P} = \int_B \mathbb{E}^{\mathcal{B}}(X_n)\, d\mathbb{P}$, on a (Beppo-Levi), $\int_B X\, d\mathbb{P} = \int_B Y\, d\mathbb{P}$ i.e. $Y = \mathbb{E}^{\mathcal{B}}(X)$ p.s.

(ii) Vu que $\underline{\lim}\, X_n = \lim_n \uparrow \inf_{i \geq n} X_i$, on a p.s., utilisant (i), $\mathbb{E}^{\mathcal{B}}(\underline{\lim}\, X_n) = \mathbb{E}^{\mathcal{B}}(\lim_n \uparrow \inf_{i \geq n} X_i) = \lim_n \uparrow \mathbb{E}^{\mathcal{B}}(\inf_{i \geq n} X_i) \leq \lim_n \uparrow \inf_{i \geq n} \mathbb{E}^{\mathcal{B}}(X_i) = \underline{\lim}\, \mathbb{E}^{\mathcal{B}}(X_n)$.

(iii) On applique (ii) aux v.a. $V + X_n$ et à $V - X_n$ qui sont positives.

(iv) On applique (iii) ∎

Espérance conditionnelle et uniforme intégrabilité

Proposition 3.2.14. *Soit* X *une variable aléatoire intégrable et* $\mathcal{B}_i$ *une famille de sous-tribus de* $\mathcal{A}$. *La famille* $X_i = \mathbb{E}(X|\mathcal{B}_i)$ *est uniformément intégrable.*

`Preuve:`
L'ensemble $A_i = \{|X_i| > a\}$ est $\mathcal{B}_i$ mesurable et d'après l'inégalité de Jensen on a $|X_i| \leq \mathbb{E}(|X||\mathcal{B}_i)$, par conséquent:

$$\mathbb{E}(|X_i|) \leq \mathbb{E}(|X|), \quad \text{et} \quad \int_{A_i} |X_i|\, d\mathbb{P} \leq \int_{A_i} |X|\, d\mathbb{P}$$

Or $\mathbb{P}(A_i) \leq \mathbb{E}(|X_i|)/a \leq \mathbb{E}(|X|)/a$. Pour $\epsilon > 0$ on sait qu'il existe $\delta > 0$ tel que $\mathbb{P}(A) \leq \delta \Rightarrow \int_A |X|\, d\mathbb{P} \leq \epsilon$. Pour ϵ donné, il suffit donc de choisir a assez grand pour que $\mathbb{P}(A_i) \leq \delta$ quel que soit i, et on aura alors $\int_{A_i} |X_i|\, d\mathbb{P} \leq \epsilon$ quel que soit i ∎

On obtient de même le résultat suivant:

Proposition 3.2.15. *Soit* X_i *une famille uniformément intégrable et* $\mathcal{B}$ *une sous-tribu de* $\mathcal{A}$. *La famille* $Y_i = \mathbb{E}(X_i|\mathcal{B})$ *est uniformément intégrable.*

`Preuve:`
L'ensemble $B_i = \{|Y_i| > a\}$ est $\mathcal{B}$ mesurable et d'après l'inégalité de Jensen on a $|Y_i| \leq \mathbb{E}(|X_i||\mathcal{B})$, par conséquent:

$$\mathbb{E}(|Y_i|) \leq \mathbb{E}(|X_i|), \quad \text{et} \quad \int_{B_i} |Y_i|\, d\mathbb{P} \leq \int_{B_i} |X_i|\, d\mathbb{P}$$

D'après (2.6.2), la famille X_i est bornée dans L^1 par une constante $C < \infty$ et donc $\mathbb{P}(B_i) \leq \mathbb{E}(|Y_i|)/a \leq \mathbb{E}(|X_i|)/a \leq C/a$. Pour $\epsilon > 0$ on sait, toujours d'après (2.6.2), que la famille X_i est équicontinue, c'est à dire qu'il existe $\delta > 0$ tel que $\mathbb{P}(B) \leq \delta \Rightarrow \int_B |X_i|\, d\mathbb{P} \leq \epsilon$. Pour ϵ donné, il suffit donc de choisir a assez grand pour que $\mathbb{P}(B_i) \leq \delta$ quel que soit i, et on aura alors $\int_{B_i} |Y_i|\, d\mathbb{P} \leq \epsilon$ quel que soit i ∎

3.3. Conditionnement par une variable aléatoire

Soit T une v.a. à valeurs $(E, \mathcal{E})$. Nous allons conditionner par T.

Soit $\sigma(T) = \sigma(T^{-1}(F),\ F \in \mathcal{E})$. On sait que (prop.1.1.4)

Proposition 3.3.1. *Soit Z une v.a. à valeurs $\mathbb{R}$ (resp. $\overline{\mathbb{R}}^+$). Z est $\sigma(T)$-mesurable ssi il existe f mesurable de $(E, \mathcal{E})$ dans $\mathbb{R}$ (resp. mesurable de $(E, \mathcal{E})$ dans $\overline{\mathbb{R}}^+$) telle que $Z = f \circ T$.*

Pour X intégrable ou positive, on définit l'espérance conditionnelle de X sachant T comme l'espérance conditionnelle de X sachant $\sigma(T)$ i.e.

$$\mathbb{E}(X|T) := \mathbb{E}(X|\sigma(T)). \tag{3.7}$$

Il résulte alors de la prop.3.2.6

Proposition 3.3.2. *Soit X une v.a. intégrable (resp. positive). Alors $Y = \mathbb{E}(X|T)$ ssi Y est de la forme $Y = h(T)$ avec $h \in \mathcal{B}_b(\mathbb{R})$ (resp. $\mathcal{B}^+(\mathbb{R})$) et*

$$\boxed{\mathbb{E}(g(T)Y) = \mathbb{E}(g(T)X)} \tag{3.8}$$

pour toute $g \in \mathcal{E}_b$, (resp. toute $g \in \mathcal{E}^+$).

La fonction h ci-dessus est unique au sens suivant. Si $\mathbb{E}(X|T) = h(T) = h'(T)$, $h(T) = h'(T)$ p.s. et donc $h = h'$ μ_T p.p., μ_T étant la loi de T. On écrit alors un peu incorrectement

$$h(t) = \mathbb{E}(X|T = t).$$

Cette écriture devient correcte si $\mathbb{P}(T = t) > 0$ et alors $h(t) = \frac{1}{\mathbb{P}(T=t)} \int_{\{T=t\}} X \, d\mathbb{P}$.

Supposons $X \in L^2$ et soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Le th.3.2.2 montre que $Y = \mathbb{E}(X|\mathcal{B})$ est la projection orthogonale de X sur $K = \{Z \in L^2,\ \text{Z a un représentant } \mathcal{B}\text{-mesurable}\} \simeq L^2(\Omega, \mathcal{B}, \mathbb{P})$. Donc $\mathbb{E}(X - Y)^2 = \inf\{E(X - Z)^2,\ Z \in L^2,\ Z\ \mathcal{B}\text{-mesurable}\}$.
Si $\mathcal{B} = \sigma(T)$, on a alors que, pour $h(T) = \mathbb{E}(X|T)$,

$$\mathbb{E}(X - h(T))^2 = \inf\{\mathbb{E}(X - g(T))^2,\ g \in L^2(\mu_T)\}$$

où μ_T est la loi de T. Donc, si $X \in L^2$, $h(T) = \mathbb{E}(X|T)$ est la meilleure approximation dans L^2 de X par une fonction de T.

3.4. Conditionnement et indépendance

Soit $(E, \mathcal{E})$ un espace mesurable.

Les principaux résultats sur l'indépendance ont été présentés section 2.2. On dit qu'une v.a. X et une tribu $\mathcal{B}$ sont indépendantes si les tribus $\sigma(X)$ et $\mathcal{B}$ sont indépendantes. Ceci équivaut à X est indépendante de toute v.a.r. $\mathcal{B}$-mesurable.

Proposition 3.4.1. *Soit X une v.a.r. positive ou intégrable indépendante de $\mathcal{B}$ sous-tribu de $\mathcal{A}$. On a $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$ p.s.*

`Preuve:`
Soit $m = \mathbb{E}(X)$. Evidemment m est $\mathcal{B}$-mesurable. On a, pour toute $Z \in \mathcal{B}^+$, $\mathbb{E}(XZ) = \mathbb{E}(X)\mathbb{E}(Z) = \mathbb{E}(mZ)$ i.e. $\mathbb{E}(X|\mathcal{B}) = m$ p.s.

Z. L'égalité $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$ n'implique évidemment pas que X soit indépendante de $\mathcal{B}$. Par contre, on a:.

Proposition 3.4.2. *Soient X une v.a. à valeurs $(E, \mathcal{E})$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Alors X est indépendante de $\mathcal{B}$ ssi*

$$\text{pour toute } f \in \mathcal{E}^+, \quad \mathbb{E}^{\mathcal{B}}(f(X)) = \mathbb{E}(f(X)) \text{ p.s.} \tag{3.9}$$

`Preuve:`
Si X est indépendante de $\mathcal{B}$, (3.9) résulte de la prop.3.4.1. Réciproquement supposons (3.9). Soient $f \in \mathcal{E}^+$ et $Z \in \mathcal{B}^+$. On a, vu les prop.3.2.5 et 3.2.9,

$$\mathbb{E}(f(X)Z) = \mathbb{E}(\mathbb{E}^{\mathcal{B}}(f(X)Z)) = \mathbb{E}(Z\mathbb{E}^{\mathcal{B}}(f(X))) = \mathbb{E}(Z\mathbb{E}(f(X))) = \mathbb{E}(Z)\mathbb{E}(f(X)).$$

et les tribus $\sigma(X)$ et $\mathcal{B}$ sont indépendantes ∎

Remarque: On vérifie facilement que, si X est indépendante de $\mathcal{B}$, (3.9) est vraie pour toute $f \in L^1(\mu_X)$.

Proposition 3.4.3. *Soient X une v.a. à valeurs $\mathbb{R}^d$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Alors X est indépendante de $\mathcal{B}$ ssi*

$$\text{pour tout } t \in \mathbb{R}^d, \ \ \mathbb{E}^{\mathcal{B}}(e^{i<t,X>}) = \mathbb{E}(e^{i<t,X>}) \text{ p.s.} \tag{3.10}$$

`Preuve:`
(i) Si X est indépendante de $\mathcal{B}$, on a (3.10) vu la prop.3.4.1 .

(ii) Supposons (3.10). Soit Z une v.a.r. $\mathcal{B}$-mesurable. On a, pour tout $t \in \mathbb{R}^p$ et tout $\tau \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}(\mathrm{e}^{i<t,X>+i\tau Z}) &= \mathbb{E}(\mathrm{e}^{i\tau Z}\mathbb{E}^{\mathcal{B}}(\mathrm{e}^{i<t,X>})) \\ &= \mathbb{E}(\mathrm{e}^{i\tau Z}\mathbb{E}(\mathrm{e}^{i<t,X>})) = \mathbb{E}(\mathrm{e}^{i<t,X>})\mathbb{E}(\mathrm{e}^{i\tau Z}) \end{aligned}$$

et donc (prop.2.7.5) X et Z sont indépendantes ∎

On peut généraliser la prop.3.4.1

Proposition 3.4.4. *Soient $\mathcal{F}$ et $\mathcal{G}$ des sous-tribus de $\mathcal{A}$ et X une v.a. intégrable. On suppose que $\mathcal{G}$ est indépendante de $\sigma(\sigma(X), \mathcal{F})$. Alors*

$$\mathbb{E}(X|\sigma(\mathcal{F}, \mathcal{G})) = \mathbb{E}(X|\mathcal{F}) \ \text{ p.s.}$$

`Preuve:`
Soit $Y = \mathbb{E}(X|\mathcal{F})$. Evidemment Y est $\sigma(\mathcal{F},\mathcal{G})$-mesurable. Soient $U \in b\mathcal{F}$ et $V \in b\mathcal{G}$, on a $\mathbb{E}(XUV) = \mathbb{E}(XU)\mathbb{E}(V) = \mathbb{E}(YU)\mathbb{E}(V) = \mathbb{E}(YUV)$. Il suffit maintenant d'appliquer la prop. 3.2.7 ∎

Corollaire 3.4.5. *Si X est intégrable ou positive et si la variable aléatoire U est indépendante du couple (T, X), on a*

$$\mathbb{E}(X|\,(T,U)) = \mathbb{E}(X|\,T) \ p.s.$$

Terminons par un résultat très utile pour le calcul d'espérances conditionnelles.

Proposition 3.4.6. *Soient $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ et T, X des v.a. à valeurs $(E,\mathcal{E})$ et $(F,\mathcal{F})$. On suppose que T est $\mathcal{B}$-mesurable et que X est indépendante de $\mathcal{B}$. Alors, pour toute $h \in (\mathcal{E}\otimes\mathcal{F})^+ \cup (\mathcal{E}\otimes\mathcal{F})_b$,*

$$\mathbb{E}(h(T,X)|\mathcal{B}) = \phi(T) \ p.s. \ où \ \phi(t) = \mathbb{E}(h(t,X)).$$

`Preuve:`
Supposons $h \geq 0$. Soit $Z \in \mathcal{B}^+$. Notons μ_X et $\mu_{T,Z}$ les lois de X et de (T,Z). On a $\phi(t) = \int h(t,x)\,d\mu_X(x)$ et

$$\mathbb{E}(Zh(T,X)) = \int zh(t,x)\,d\mu_{T,Z}(t,z)\,d\mu_X(x) =$$
$$\int z[\int h(t,x)\,d\mu_X(x)]\,d\mu_{T,Z}(t,z) = \int z\phi(t)\,d\mu_{T,Z}(t,z) = \mathbb{E}(Z\phi(T))$$

d'où $\phi(T) = \mathbb{E}(h(T,X)|\mathcal{B})$ p.s. ∎

Corollaire 3.4.7. *Avec les notations de la Proposition 3.4.6, si T et X sont des variables aléatoires indépendantes, on a*

$$\mathbb{E}(h(T,X)|\,T=t) = \mathbb{E}(h(t,X)) \quad p.s.$$

3.5. Lois conditionnelles

Commençons par un exemple élémentaire. Soient X, Y des v.a. entières indépendantes de loi de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. Soit $T = X+Y$. On sait que que $T \sim \mathcal{P}(\lambda+\mu)$ et l'on a, pour $0 \leq k \leq t$,

$$\begin{aligned}
\mathbb{P}(X=k|\,T=t) &= \frac{\mathbb{P}(X=k,T=t)}{\mathbb{P}(T=t)} = \frac{\mathbb{P}(X=k)\mathbb{P}(Y=t-k)}{\mathbb{P}(T=t)} \\
&= \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\mathrm{e}^{-\mu}\frac{\mu^{t-k}}{(t-k)!}\mathrm{e}^{\lambda+\mu}\frac{t!}{(\lambda+\mu)^t} = \frac{t!}{k!(t-k)!}\frac{\lambda^k\mu^{t-k}}{\lambda+\mu)^t} \\
&= C_t^k p^k(1-p)^{(t-k)} = N_t(k) \text{ où } p = \frac{\lambda}{\lambda+\mu}.
\end{aligned}$$

On dira que $(N_t(k),\ 0 \le k \le t)$ est la loi conditionnelle de X sachant que $T = t$. Ici il s'agit de la loi binomiale $B(t, \frac{\lambda}{\lambda+\mu}.)$. On a alors, pour toute f positive,

$$\mathbb{E}(f(X)|\,T = t) = \sum_k f(k)N_t(k).$$

Plus généralement, soient T, X des v.a. à valeurs $(E, \mathcal{E})$ et $(F, \mathcal{F})$. On sait que μ est la loi de X ssi, pour tout $f \in \mathcal{F}^+$, $\mathbb{E}(f(X)) = \int f(x)\,d\mu(x)$. On appellera donc la loi conditionnelle de X sachant que $T = t$, une famille de probabilités sur $(F, \mathcal{F})$, soit $(N_t(dx),\ t \in E)$, telle que, pour toute $f \in \mathcal{F}^+$,

$$\mathbb{E}(f(X)|\,T = t) = \int f(x)N_t(dx).$$

Pour être tout à fait rigoureux, il faut préciser les mesurabilités.

Définition 3.5.1. *On appelle probabilité de transition de E dans F une famille de probabilités sur $(F, \mathcal{F})$, soit $(N_t(dx),\ t \in E)$, telle que, pour tout $A \in F$, $t \mapsto N_t(A)$ soit $\mathcal{E}$-mesurable. On adopte souvent la notation $N(t, dx)$ plutôt que $N_t(dx)$. Pour toute fonction f, $\mathcal{F}$ mesurable, positive (resp. bornée) on définit*

$$Nf(t) = \int_F f(x)\ N(t, dx)$$

qui est une une fonction $\mathcal{E}$ mesurable positive (resp. bornée).

Définition 3.5.2. *Une probabilité de transition, $(N(t, dx), t \in E)$, de E dans F est la loi conditionnelle de X sachant que $T = t$ si, pour toute $f \in \mathcal{F}^+$,*

$$\mathbb{E}(f(X)|\,T) = Nf(T) \ \ p.s., \ \ ou\ encore\ \ \mathbb{E}(f(X)|\,T = t) = Nf(t)\ \ \mu_T p.s., \tag{3.11}$$

où $\mu_T =$ loi de T.

Il existe un cas ou le calcul de cette loi conditionnelle est immédiat:

Proposition 3.5.3. *Si X et T sont des variables aléatoires indépendantes , alors pour toute fonction $\varphi \in (\mathcal{E} \otimes \mathcal{F})$ la loi conditionnelle de $\varphi(T, X)$ sachant que $T = t$ est identique à la loi de la variable aléatoire $\varphi(t, X)$.*

`Preuve:`
Il suffit d'appliquer le Corollaire 3.4.7. ∎

La prop.3.3.2 entraîne que

Proposition 3.5.4. *Une probabilité de transition de E dans F, $(N(t, dx), t \in E)$, est la loi conditionnelle de X sachant que $T = t$ ssi, pour toute $f \in \mathcal{F}^+$ et toute $g \in \mathcal{E}^+$,*

$$\boxed{\mathbb{E}[g(T)f(X)] = \int g(t)Nf(t)\,d\mu_T(t)\ \ où\ \mu_T =\ loi\ de\ T} \tag{3.12}$$

La loi conditionnelle de X sachant que $T = t$ est unique au sens suivant. Si $N(t,dx)$ et $N'(t,dx)$ vérifient (3.11), on a, pour tout $A \in \mathcal{F}$, $N(T,A) = N'(T,A)$ p.s. i.e. $N(t,A) = N'(t,A)$ μ_T p.p.
La formule (3.12) montre que si on connaît la loi de T et la loi conditionnelle de X sachant que $T = t$, on peut reconstituer la loi du couple (X,T). Plus précisément,

Proposition 3.5.5. *Soient T, X des v.a. à valeurs $(E,\mathcal{E})$ et $(F,\mathcal{F})$, μ_T la loi de T, $N(t,dx)$ la loi conditionnelle de X sachant que $T = t$. On a , pour toute $h \in (\mathcal{E} \otimes \mathcal{F})^+ \cup (\mathcal{E} \otimes \mathcal{F})_b$,*

$$\mathbb{E}[h(T,X)] = \int [\int h(t,x)\ N(t,dx)]\, d\mu_T(t).$$

`Preuve:`
Soient $\mu_1(C) = \mathbb{P}((T,X) \in C)$ et $\mu_2(C) = \int[\int \mathbb{1}_{\{C\}}(t,x)N(t,dx)]\, d\mu_T(t)$. μ_1 et μ_2 sont deux probabilités sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$ (μ_1 est la loi de (T,X)). On a $\mathbb{E}[h(T,X)] = \int h(t,x)\, d\mu_1(t,x)$ et $\int[\int h(t,x)N(t,dx)]\, d\mu_T(t) = \int h(t,x)\, d\mu_2(t,x)$. Vu la prop.3.5.4, $\mu_1(A \times B) = \mu_2(A \times B)$ pour tout $A \in \mathcal{E}$ et $B \in \mathcal{F}$ et donc (prop.1.2.3) $\mu_1 = \mu_2$ et le résultat cherché ∎

Les problèmes classiques de conditionnement se traitent grâce à

Proposition 3.5.6. *Soient T, X des v.a. à valeurs $(E,\mathcal{E})$ et $(F,\mathcal{F})$, α, β des mesures σ-finies sur $(E,\mathcal{E})$ et $(F,\mathcal{F})$. On suppose que (T,X) a une densité $h(t,x)$ par rapport à $\alpha \otimes \beta$. On pose $\phi(t) = \int h(t,x)\, d\beta(x)$ et*

$$h(x/t) = \frac{h(t,x)}{\phi(t)} \text{ si } \phi(t) \neq 0, \quad h(x/t) = \text{densité arbitraire si } \phi(t) = 0.$$

Alors $N(t,dx) = h(x/t)\ d\beta(x)$ est la loi conditionnelle de X sachant que $T = t$.

`Preuve:`
Notons d'abord que ϕ est la densité de T par rapport à α. Soit $B = \{t,\ \phi(t) = 0\}$. On a $\int_{B\times F} h(t,x)\, d\alpha(t)d\beta(x) = \int_B \phi(t)\, d\alpha(t) = 0$ et $h(t,x) = 0$ sur $B \times F$ $\alpha \otimes \beta$ p.p. On en déduit que $h(t,x) = \phi(t)h(x/t)$ $\alpha \otimes \beta$ p.p. Soient $f \in \mathcal{F}^+$ et $g \in \mathcal{E}^+$, on a

$$\begin{aligned}\mathbb{E}[g(T)f(X)] &= \int g(t)f(x)h(t,x)\, d\alpha(t)d\beta(x) = \int g(t)f(x)\phi(t)h(x/t)\, d\alpha(t)d\beta(x) \\ &= \int g(t)[\int f(x)h(x/t)\, d\beta(x)]\phi(t)\, d\alpha(t)\end{aligned}$$

et $N(t,dx) = h(x/t)\ d\beta(x)$ par la prop.3.5.4 ∎

La fonction $h(x/t)$ s'appelle la densité conditionnelle de X sachant que $T = t$. On a donc

$$h(x/t) = \frac{\text{densité de } (T,X)}{\text{densité de } T},$$

ou, de façon heuristique,

$$h(x/t) = \mathbb{P}(X \in dx |\, T \in dt) = \frac{\mathbb{P}(T \in dt,\ X \in dx)}{\mathbb{P}(T \in dt)}.$$

Ceci permet de calculer des espérances conditionnelles puisque, si $F = \mathbb{R}$, $\mathbb{E}(X|T = t) = \int x n(t, dx)$. On a donc dans ce cas

$$\mathbb{E}(X|T = t) = \frac{\int x h(t,x)\, d\beta(x)}{\int h(t,x)\, d\beta(x)}. \tag{3.13}$$

3.6. Annexe

3.6.1. Un exemple

Soient X et Y des v.a. positives, indépendantes, de même loi de densité e^{-x} par rapport à λ_+ mesure de lebesgue sur $\mathbb{R}^+$. Soient $T = X+Y$ et $U = \max(X,Y)$. On veut calculer les lois conditionnelles de X sachant que $T = t$ et de X sachant que $U = u$.

(i) Il est facile de calculer la densité de (T,X) par rapport à $\lambda_+ \otimes \lambda_+$. On a, pour $g \geq 0$ arbitraire,

$$\mathbb{E}(g(T,X)) = \int_0^\infty \int_0^\infty g(x+y, x)\mathrm{e}^{-(x+y)}\, dxdy = \int_0^\infty \int_0^t g(t,x)\mathrm{e}^{-t}\, dtdx$$

et (T,X) a pour densité $h(t,x) = \mathrm{e}^{-t}\mathbb{1}_{\{0<x<t\}}$. Alors la densité de T est $\phi(t) = \int h(t,x)\, dx = t\mathrm{e}^{-t}$. On a donc

$$h(x/t) = \frac{h(t,x)}{\phi(t)} = \frac{1}{t}\mathbb{1}_{\{]0,t[\}}(x).$$

La loi conditionnelle de X sachant que $T = t$ est donc la loi uniforme sur $]0,t[$. On a en particulier $\mathbb{E}(X|T) = \frac{T}{2}$ p.s.

(ii) Si on conditionne par U, la situation est plus délicate. En effet $\mathbb{P}(X = U) = \frac{1}{2}$ et le couple (U,X) n'a pas de densité par rapport à une mesure produit. Il faut utiliser la formule générale (3.12). Calculons d'abord la loi de U. On a $\mathbb{P}(U \leq u) = \mathbb{P}(X \leq u, Y \leq u) = \mathbb{P}(X \leq u)\mathbb{P}(Y \leq u) = (1-\mathrm{e}^{-u})^2$ d'où U a pour densité $\phi(u) = 2\mathrm{e}^{-u}(1-\mathrm{e}^{-u})$. Pour f, g positives, on a

$$\begin{aligned}
\mathbb{E}(g(U)f(X)) &= \mathbb{E}(g(U)f(X)\mathbb{1}_{\{X\leq Y\}}) + \mathbb{E}(g(U)f(X)\mathbb{1}_{\{X>Y\}}) \\
&= \mathbb{E}(g(Y)f(X)\mathbb{1}_{\{X\leq Y\}}) + \mathbb{E}(g(X)f(X)\mathbb{1}_{\{X>Y\}}) \\
&= \int_0^\infty \int_0^\infty g(y)f(x)\mathbb{1}_{\{x\leq y\}}\mathrm{e}^{-(x+y)}dxdy \\
&+ \int_0^\infty \int_0^\infty g(x)f(x)\mathbb{1}_{\{x>y\}}\mathrm{e}^{-(x+y)}dxdy \\
&= \int_0^\infty g(y)e^{-y}\int_0^y f(x)\mathrm{e}^{-x}\, dx\, dy + \int_0^\infty g(x)f(x)e^{-x}(1-\mathrm{e}^{-x})\, dx \\
&= \int_0^\infty g(u)\frac{1}{2(1-\mathrm{e}^{-u})}[\int_0^u f(x)\mathrm{e}^{-x}\, dx]\phi(u)\, du + \int_0^\infty g(u)\frac{1}{2}f(u)\phi(u)\, du \\
&= \int_0^\infty g(u)\,[\frac{1}{2(1-\mathrm{e}^{-u})}\int_0^u f(x)\mathrm{e}^{-x}\, dx + \frac{1}{2}f(u)]\,\phi(u)\, du
\end{aligned}$$

donc la loi conditionnelle de X sachant que $U = u$ est

$$N(u, dx) = \frac{1}{2(1 - e^{-u})} e^{-x} \mathbb{1}_{\{]0,u[\}}(x)\, dx + \frac{1}{2}\delta_u(dx)$$

où δ_u est la mesure de Dirac du point u et on a

$$\mathbb{E}(X|\, U) = \frac{1}{2(1 - e^{-U})} \int_0^U x e^{-x}\, dx + \frac{U}{2} = \frac{1 + U - (1 + 2U)e^{-U}}{2(1 - e^{-U})} \text{ p.s.}$$

3.6.2. Le cas gaussien

On considère des v.a. T et X à valeurs $\mathbb{R}^p$ et $\mathbb{R}^q$ et on suppose que le couple (T, X) est gaussien de matrice de covariance

$$K(T, X) = \begin{pmatrix} K(T) & \Sigma_{12} \\ \Sigma_{21} & K(X) \end{pmatrix}. \tag{3.14}$$

On suppose que T est non dégénérée i.e. que det(K(T)) $\neq 0$. On s'intéresse à la loi conditionnelle de X sachant que $T = t$. On pose $\tilde{T} = T - \mathbb{E}(T)$, $\tilde{X} = X - \mathbb{E}(X)$. On a donc $K(T) = \mathbb{E}(\tilde{T}\tilde{T}')$, $K(X) = \mathbb{E}(\tilde{X}\tilde{X}')$, $\Sigma_{12} = \mathbb{E}(\tilde{T}\tilde{X}')$, $\Sigma_{21} = \mathbb{E}(\tilde{X}\tilde{T}') = \Sigma_{12}'$. On cherche à écrire X sous la forme $X = AT + Z$ avec T et Z indépendantes ce qui revient à chercher une matrice A telle que $X - AT$ et T soient indépendantes. Le couple $(X - AT, T)$ étant gaussien comme fonction linéaire de (T, X), $X - AT$ et T sont indépendantes (th.2.7.14) ssi, pour tous i, j, $\text{Cov}((X - AT)_i, T_j) = 0$ ce qui s'écrit $\mathbb{E}((\tilde{X} - A\tilde{T})\tilde{T}') = 0$ soit encore $\Sigma_{21} - AK(T) = 0$.
On a donc, pour $A = \Sigma_{21}K^{-1}(T)$, $X = AT + Z$ avec T et $Z = X - AT$ indépendantes. Evidemment la v.a. Z est gaussienne avec $\mathbb{E}(Z) = \mathbb{E}(X) - A\mathbb{E}(T)$ et $K(Z) = \mathbb{E}((\tilde{X} - A\tilde{T})(\tilde{X} - A\tilde{T})') = K(X) - \Sigma_{21}K^{-1}(T)\Sigma_{12}$. On en déduit, (cor.3.5.3), que la loi conditionnelle de X sachant que $T = t$ est la loi de $At + Z$. On a établi:

Proposition 3.6.1. *Soit (T, X) un vecteur gaussien de matrice de covariance donnée par (3.14) avec det(K(T)) $\neq 0$. La loi conditionnelle de X sachant que $T = t$ est une loi gaussienne de moyenne $\mathbb{E}(X) + \Sigma_{21}K^{-1}(T)(t - \mathbb{E}(T))$ et de matrice de covariance $K(X) - \Sigma_{21}K^{-1}(T)\Sigma_{12}$. En particulier*

$$\mathbb{E}(X \,|\, T) = \mathbb{E}(X) + \Sigma_{21}K^{-1}(T)(T - \mathbb{E}(T)) \text{ p.s.}$$

Chapitre 4

Chaînes de Markov

4.1. Processus aléatoires

Généralités

Définition 4.1.1. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité. On appelle processus aléatoire à valeurs* $(E, \mathcal{E})$ *un terme* $X = (\Omega, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P})$ *où pour tout* n, X_n *est une v.a. à valeurs* $(E, \mathcal{E})$.

L'application $n \mapsto X_n(\omega)$ s'appelle la trajectoire associée à la réalisation ω.

Posons $\mathcal{F}_n^0 = \sigma(X_0, X_1, \dots, X_n)$. On sait (prop.3.3.1) qu'alors une v.a.r. Z est $\mathcal{F}_n^0$-mesurable ssi $Z = \phi(X_0, X_1, \dots, X_n)$ avec $\phi \in \mathcal{E}$. En particulier un événement A appartient à $\mathcal{F}_n^0$ ssi

$$A = \{\omega,\ \phi(X_0(\omega), X_1(\omega), \dots, X_n(\omega)) = 1\}$$

avec $\phi \in \mathcal{E}$. Les événements de $\mathcal{F}_n^0$ sont donc les événements dont on sait s'ils ont eu lieu ou non si on connaît les valeurs prises par $X_0, X_1, \dots, X_n$. Cependant, dans de nombreux cas, ce qu'on connaît à l'instant n ne se résume pas aux seules valeurs de $X_0, X_1, \dots, X_n$, on peut très bien connaître les valeurs prises par d'autres processus jusqu'à l'instant n. C'est pourquoi on introduit,

Définition 4.1.2. *On appelle espace de probabilité filtré un terme* $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$ *où* $(\Omega, \mathcal{F}, \mathbb{P})$ *est un espace de probabilité et où* $\mathcal{F}_n$ *est une famille croissante de sous-tribus de* $\mathcal{F}$.

La tribu $\mathcal{F}_n$ s'appelle la tribu des événements antérieurs à n.

Définition 4.1.3. *On appelle processus aléatoire adapté à valeurs* $(E, \mathcal{E})$ *un terme* $X = (\Omega, \mathcal{F}, \mathcal{F}_n, (X_n)_{n\geq 0}, \mathbb{P})$ *où* $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$ *est un espace de probabilité filtré et où, pour tout* n, X_n *est une v.a.* $\mathcal{F}_n$*-mesurable à valeurs* $(E, \mathcal{E})$.

Un processus au sens de la def.4.1.1 est un processus adapté au sens de la def.4.1.3 si on choisit $\mathcal{F}_n = \mathcal{F}_n^0 = \sigma(X_0, X_1, \dots, X_n)$.

4.1.1. Processus canoniques

Soit X un processus à valeurs $(E,\mathcal{E})$. On note μ_n la loi de $(X_0, X_1, \ldots, X_n)$. C'est une probabilité sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$. Si on note $\pi_{n+1,n}$ la projection canonique de E^{n+1} sur E^n i.e. l'appliquation $(x_0, \ldots, x_{n-1}, x_n) \to (x_0, \ldots, x_{n-1})$, on a

$$\pi_{n+1,n}(\mu_n) = \mu_{n-1}.$$

En vertu de la prop.1.2.3, ceci est équivalent à, pour tout $A_k \in \mathcal{E}$,

$$\mu_{n-1}(A_0 \times A_1 \times \ldots \times A_{n-1}) = \mu_n(A_0 \times A_1 \times \ldots \times A_{n-1} \times E). \qquad (4.1)$$

Les probabilités $(\mu_n)_{n\geq 0}$ s'appellent les répartitions finies du processus X.
Réciproquement, si on se donne des probabilités μ_n sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ vérifiant (4.1), se pose la question de savoir s'il existe un processus ayant pour répartitions finies les μ_n. On introduit l'espace canonique,

$$\Omega = E^{\mathbb{N}},\ \omega = (\omega_n)_{n\geq 0},\ X_n(\omega) = \omega_n,\ \mathcal{F}_n = \sigma(X_k,\ k \leq n),\ \mathcal{F} = \sigma(X_k,\ k \geq 0). \qquad (4.2)$$

Soit $A \in \mathcal{F}_n$, A est de la forme $A = B \times E \ldots \times E \times \ldots$ avec $B \in \mathcal{E}^{\otimes(n+1)}$. On définit alors une probabilité $\mathbb{P}_n$ sur $(\Omega, \mathcal{F}_n)$ en posant,

$$\mathbb{P}_n(A) = \mu_n(B)$$

puis une fonction d'ensembles sur $\cup_n \mathcal{F}_n$ par

$$\mathbb{P}(A) = \mathbb{P}_n(A) \text{ si } A \in \mathcal{F}_n.$$

Il s'agit de prolonger $\mathbb{P}$ en une probabilité sur $\sigma(\cup_n \mathcal{F}_n)$. Remarquons que $\cup_n \mathcal{F}_n$ étant stable par intersection finie, ce prolongement sera unique (prop.1.2.3). L'existence de ce prolongement a été montrée par Kolmogorov et on a:

Théorème 4.1.4. *Soit $(\mu_n)_{n\geq 0}$ une famille de probabilités sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ vérifiant (4.1). Il existe une unique probabilité $\mathbb{P}$ sur l'espace canonique $(\Omega, \mathcal{F})$ défini par (4.2) telle $(\Omega, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P})$ soit un processus de répartitions finies $(\mu_n)_{n\geq 0}$.*

Exemple 1. Soient $\nu_0, \ldots, \nu_n \ldots$ une suite de probabilités sur $(E,\mathcal{E})$. On veut construire un modèle pour une suite de v.a. indépendantes de lois $\nu_0, \ldots, \nu_n, \ldots$. On définit μ_n sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ par $\mu_n = \nu_0 \otimes \ldots \otimes \nu_n$ et on applique le th.4.1.4. On obtient une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ telle que les X_n définies par (4.2) soient une suite de v.a. indépendantes de loi $\nu_0, \ldots, \nu_n, \ldots$.

Exemple 2. Cet exemple est à la base de la construction des chaînes de Markov que l'on étudiera plus loin. On considère un ensemble E dénombrable muni d'une probabilité μ et d'une matrice de transition $Q(x,y)$, $x, y \in E$, c'est à dire une matrice à termes positifs telle que pour tous $x, y \in E$,

$$Q(x,y) \geq 0, \qquad \sum_{y \in E} Q(x,y) = 1.$$

On définit μ_n sur E^{n+1} par $\mu_n(x_0, x_1, \ldots, x_n) = \mu(x_0)Q(x_0, x_1)\ldots Q(x_{n-1}, x_n)$ et on applique le th.4.1.4. On obtient une probabilité $\mathbb{P}_\mu$ sur $(\Omega, \mathcal{F})$ telle que les vecteurs $(X_0, X_1, \ldots, X_n)$ définis par (4.2) aient pour loi μ_n.

Processus équivalents
Soit $Y = (W, \mathcal{G}, (Y_n)_{n\geq 0}, \mathbb{Q})$ un processus à valeurs $(E, \mathcal{E})$. On définit une application mesurable (vu la définition de $\mathcal{F}$)

$$\Phi: \ (W, \mathcal{G}) \to (\Omega, \mathcal{F}), \ \ w \mapsto (Y_n(w))_{n\in\mathbb{N}}.$$

Soit $\mathbb{P}$ l'image de $\mathbb{Q}$ par Φ c'est à dire la probabilité définie sur $(\Omega, \mathcal{F})$ par $\mathbb{P}(A) = \mathbb{Q}(\Phi^{-1}(A))$, $A \in \mathcal{F}$. Cette probabilité $\mathbb{P}$ s'appelle la loi du processus Y. On a, pour tout $B_0, \ldots, B_n \in \mathcal{E}$,

$$\mathbb{P}(B_0 \times \ldots \times B_n \times E \times \ldots) = \mathbb{P}(X_0 \in B_0, \ldots, X_n \in B_n) = \mathbb{Q}(Y_0 \in B_0, \ldots, Y_n \in B_n)$$

ce qui implique que $\mathbb{P}$ est la probabilité sur l'espace canonique associée par le th.4.1.4 aux répartitions finies de Y. Le processus $(\Omega, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P})$ s'appelle alors la réalisation canonique de Y.
On dit que deux processus sont équivalents s'ils ont même loi. Donc deux processus ayant mêmes répartitions finies sont équivalents et un processus est toujours équivalent à sa réalisation canonique.

4.1.2. Temps d'arrêt

Soient $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$ un espace de probabilité filtré et $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, \, n \geq 0)$.

Définition 4.1.5. *Soit T une application de Ω dans $\overline{\mathbb{N}}$*

(i) *On dit que T est un temps d'arrêt (de la filtration $\mathcal{F}_n$) si pour tout entier $n \geq 0$, $\{T \leq n\} \in \mathcal{F}_n$.*

(ii) *On appelle tribu des événements antérieurs à T et on note $\mathcal{F}_T$ la tribu*

$$\mathcal{F}_T = \{\, A \in \mathcal{F}_\infty, \ \text{pour tout } n \geq 0, \ A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

Remarque 1. On vérifie immédiatement que $\mathcal{F}_T$ est une tribu et que, si T est constant et égal à n alors T est un temps d'arrêt et $\mathcal{F}_T = \mathcal{F}_n$.

Remarque 2. Vu les formules

$$\{T \leq n\} = \cup_{k=0}^n \{T = k\}, \ \ \{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$$

on peut remplacer dans (i) et (ii) $\{T \leq n\}$ et $A \cap \{T \leq n\}$ par $\{T = n\}$ et $A \cap \{T = n\}$.

On note $\mathcal{T}$ l'ensemble des temps d'arrêt de la filtration $\mathcal{F}_n$. On a les propriétés suivantes,

Proposition 4.1.6. *Soit* $T_1, T_2 \in \mathcal{T}$ *alors:*

(i) $T_1 + T_2$, $T_1 \wedge T_2$ *sont dans* $\mathcal{T}$,

(ii) *si* $T_1 \leq T_2$, *alors* $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$.

`Preuve:`
(i) D'une part $\{T_1 + T_2 = n\} = \cup_{k=0}^{n} \{T_1 = k\} \cap \{T_2 = n - k\} \in \mathcal{F}_n$. D'autre part $\{T_1 \wedge T_2 = n\} = \{T_1 = n\} \cap \{T_2 \geq n\} \cup \{T_1 \geq n\} \cap \{T_2 = n\} \in \mathcal{F}_n$.

(ii) Soit $A \in \mathcal{F}_{T_1}$. On a $A \cap \{T_2 = n\} = \cup_{k=0}^{n} A \cap \{T_1 = k\} \cap \{T_2 = n\} \in \mathcal{F}_n$ et donc $A \in \mathcal{F}_{T_2}$ ∎

Soit X un processus adapté, on veut définir la position du processus à l'instant T i.e. $X_{T(\omega)}(\omega)$. Il y a un problème lorsque $T(\omega) = +\infty$. On se donne une v.a. X_∞, $\mathcal{F}_\infty$-mesurable et on pose

$$X_T = X_n \text{ sur } \{T = n\} \;\; n \in \overline{\mathbb{N}}.$$

Proposition 4.1.7. *Soit* T *un temps d'arrêt . Les v.a.* T *et* X_T *sont* $\mathcal{F}_T$*-mesurables.*

`Preuve:`
Pour T, ceci résulte de la définition. Pour X_T, on a $\{X_T \in A\} \cap \{T = n\} = \{X_n \in A\} \cap \{T = n\} \in \mathcal{F}_n$ ∎

Exemples de temps d'arrêt
Soit X un processus aléatoire adapté, à valeurs dans $(E, \mathcal{E})$.

Définition 4.1.8. *Pour* $A \in \mathcal{E}$, *on définit (avec la convention* $\inf \emptyset = +\infty$*):*

- *Le temps d'entrée du processus dans* A *soit* $T_A = \inf(n \geq 0,\ X_n \in A)$
- *Le temps de retour du processus dans* A *soit* $S_A = \inf(n \geq 1,\ X_n \in A)$
- *Les temps de retour successifs du processus dans* A *soit* $S_A^0 = 0$, *puis pour* $k \geq 0$, $S_A^{k+1} = \inf(n > S_A^k,\ X_n \in A)$.
- *Le nombre de visites du processus à* A *soit* $N_A = \sum_{n=0}^{\infty} \mathbb{1}_{\{A\}} \circ X_n$

Proposition 4.1.9. T_A, S_A, S_A^n *sont des temps d'arrêt et* $N_A = \mathbb{1}_{\{A\}} \circ X_0 + \sum_{n=1}^{\infty} \mathbb{1}_{\{S_A^n < +\infty\}}$.

`Preuve:`
On peut écrire $\{T_A = n\} = \{X_0 \notin A, X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} \in \mathcal{F}_n$. On montre de même que $S_A = S_A^1$ est un temps d'arrêt puis, par récurrence, on obtient le même résultat pour S_A^n en écrivant:

$$\{S_A^{n+1} = p\} = \cup_{k=1}^{p-1} \{(S_A^n = k), X_{k+1} \notin A, \ldots, X_{p-1} \notin A, X_p \in A\}$$

Z. Lorsque T est un temps d'arrêt p.s. fini et X un processus adapté, on peut définir sans problème la variable aléatoire X_T. Il se peut très bien que tous les X_n soient

intégrables (et même que $\sup_n \mathbb{E}(|X_n|) < \infty$) sans que X_T le soit. Par exemple, sur $\Omega = \mathbb{N}^*$ on considère la filtration $\mathcal{F}_n = \sigma(\{1\}, \{2\}, \ldots, \{n\}, [n+1, \ldots \infty[)$, le processus adapté X_n défini par $X_n(i) = n\mathbb{1}_{\{n\}}(i)$ et le temps d'arrêt fini $T(i) = i$. On a alors $X_T(i) = i$. Si l'on choisit la probabilité $\mathbb{P}$ telle que $\mathbb{P}(n) = c/n^2$ on a $\mathbb{E}(X_T) = \infty$ alors que $\mathbb{E}(X_n) = c/n$. Bien sûr, une telle situation ne peut se produire s'il existe une variable aléatoire intégrable U telle que $|X_n| \leq U$ puisqu'alors $|X_T| \leq U$. Le résultat suivant fournit cependant une formule très utile dans le cas particulier d'un temps d'arrêt intégrable associé à une marche aléatoire.

Identité de Wald.

Proposition 4.1.10. *Soient $(Y_n)_{n\geq 1}$ une suite de v.a.r. indépendantes, intégrables, de même loi et T un temps d'arrêt intégrable. On pose $X_0 = 0$ et $X_n = Y_1 + Y_2 + \ldots + Y_n$ pour $n \geq 1$. Alors X_T est intégrable et $\mathbb{E}(X_T) = \mathbb{E}(Y_1)\mathbb{E}(T)$.*

`Preuve:`
Puisque T est intégrable, $P(T < +\infty) = 1$. De plus, on peut supposer que les Y_n sont des variables aléatoires positives, il suffira dans le cas général d'appliquer le résultat obtenu aux suites Y_n^+ et Y_n^-. On a alors:

$$\begin{aligned}
X_T &= \sum_{n=1}^{T} Y_n = \sum_{n\geq 0} Y_{n+1}\mathbb{1}_{\{T\geq n+1\}} \\
\mathbb{E}(X_T) &= \sum_{n\geq 0} \mathbb{E}(Y_{n+1}\mathbb{1}_{\{T\geq n+1\}}) = \sum_{n\geq 0} \mathbb{E}(\mathbb{E}(Y_{n+1}\mathbb{1}_{\{T\geq n+1\}}|\mathcal{F}_n)) \\
&= \sum_{n\geq 0} \mathbb{E}(\mathbb{1}_{\{T\geq n+1\}}\mathbb{E}(Y_{n+1}|\mathcal{F}_n)) = \mathbb{E}(Y_1)\sum_{n\geq 0} \mathbb{E}(\mathbb{1}_{\{T\geq n+1\}}) \\
&= \mathbb{E}(Y_1)\mathbb{E}(T)
\end{aligned}$$

la troisième ligne ci dessus étant obtenue en utilisant le fait que $\{T \geq n+1\} \in \mathcal{F}_n$ et que la variable aléatoire Y_{n+1} est indépendante de $\mathcal{F}_n$ ∎

Dans la suite de ce chapitre E désigne un espace fini ou dénombrable.

4.2. Matrices de transition

On note $\mathcal{E}^+$ l'ensemble des applications de E dans $\overline{\mathbb{R}}^+$, $\mathcal{E}_b$ l'ensemble des applications bornées de E dans $\mathbb{R}$, $\mathcal{M}^+$ l'ensemble des mesures (positives) sur E, $\mathcal{M}_1 = \{\mu \in \mathcal{M}^+,\ \sum_{x\in E}\mu(x) = 1\}$. Evidemment E étant dénombrable, $\mu \in \mathcal{M}^+$ est déterminée par les $\mu(x),\ x \in E$.

Définition 4.2.1. *Une matrice positive M est une famille $(M(x,y),\ x,y \in E)$ où, pour tous x, y, $M(x,y) \in \overline{\mathbb{R}}^+$. Pour $A \subset E$, on pose $M(x,A) = \sum_{y\in A} M(x,y)$*

On rappelle que, dans $\overline{\mathbb{R}}^+$, $(+\infty)\times 0 = 0\times(+\infty) = 0$. Etant données deux matrices positives M et N, on note MN la matrice produit définie par

$$MN(x,y) = \sum_{z\in E} M(x,z)N(z,y)$$

Comme pour le produit ordinaire des matrices, celui-ci est associatif mais non commutatif. On note $M^0 = I$ où $I(x,y) = \mathbb{1}_{\{x=y\}}$ et, pour $n \geq 1$, $M^n = MM^{n-1} = M^{n-1}M$.

Définition 4.2.2. *Soit $f \in \mathcal{E}^+$ et $\mu \in \mathcal{M}^+$ on pose:*

$$Mf(x) = \sum_{y\in E} M(x,y)f(y), \qquad \mu M(y) = \sum_{x\in E} \mu(x)M(x,y)$$

alors $f \mapsto Mf$ est un opérateur "linéaire" de $\mathcal{E}^+$ dans $\mathcal{E}^+$ et la composition de ces opérateurs correspond aux produits des matrices i.e. l'on a $M(Nf)(x) = (MN)f(x)$, de plus $M(x,A) = M\mathbb{1}_{\{A\}}(x)$. De même $\mu \mapsto \mu M$ est un opérateur "linéaire" de $\mathcal{M}^+$ dans $\mathcal{M}^+$ et on a $(\mu M)N = \mu(MN)$, $(\mu M)f = \mu(Mf)$.

Les matrices positives telles que, pour tout $x \in E$, $A \mapsto M(x,A)$ soit une probabilité sur E seront appelées matrices de transition. D'où

Définition 4.2.3. *On appelle matrice de transition sur E une famille $(Q(x,y),\ x,y \in E)$ telle que, pour tous $x,y \in E$,*

$$Q(x,y) \geq 0, \qquad \sum_{y\in E} Q(x,y) = 1.$$

On vérifie immédiatement qu'une matrice positive Q est une matrice de transition ssi $Q1 = 1$, que le produit de deux matrices de transition est une matrice de transition, que, si Q est une matrice de transition, Q opère sur b$\mathcal{E}$ et, que, si μ une probabilité, alors μQ est une probabilité.

Si E est fini, E peut s'identifier à $\{1,2,\ldots,n\}$. Alors on identifie $f \in \mathcal{E}_b$ au vecteur colonne de composantes $(f(1), f(2), \ldots, f(n))$ et Qf est le vecteur colonne produit de la matrice Q par le vecteur colonne f. De même, on identifie $\mu \in \mathcal{M}^+$ au vecteur ligne de composantes $(\mu(1), \mu(2), \ldots, \mu(n))$ et μQ est le vecteur ligne produit du vecteur ligne μ par la matrice Q.

4.3. Suites markoviennes

Nous allons préciser la définition d'une chaîne de Markov.

Définition 4.3.1. *Soit Q une matrice de transition sur E. Un processus $X = (\Omega, \mathcal{F}, \mathcal{F}_n, X_n, \mathbb{P})$ à valeurs dans E est une chaîne de Markov (homogène) de matrice de transition Q et relativement à la filtration $\mathcal{F}_n$ si, pour tout $f \in \mathcal{E}_b$*

$$\mathbb{E}(f(X_{n+1})|\,\mathcal{F}_n) = Qf(X_n) \ \ p.s. \tag{4.3}$$

La loi de X_0 s'appelle la loi initiale de la chaîne.

A partir de maintenant nous ne considérerons que des chaînes de Markov homogènes et nous omettrons le mot "homogène". Considérons les tribus $\mathcal{F}_n^0 = \sigma(X_0, \ldots, X_n)$. De l'inclusion $\mathcal{F}_n^0 \subset \mathcal{F}_n$ il résulte immédiatement que

$$\mathbb{E}(f(X_{n+1})|\,\mathcal{F}_n^0) = \mathbb{E}\left((\mathbb{E}(f(X_{n+1})|\,\mathcal{F}_n))|\mathcal{F}_n^0)\right) = Qf(X_n)$$

Par conséquent la chaîne est aussi de Markov relativement à la filtration $\mathcal{F}_n^0$. C'est donc cette filtration qui sera choisie si aucune n'est précisée, mais quelquefois, ce n'est pas celle qui est la mieux adaptée pour prouver qu'une suite est markovienne. On peut aussi remarquer que la définition est équivalente à:

$$\mathbb{P}((X_{n+1} = y)|\,\mathcal{F}_n) = Q(X_n, y) = \mathbb{P}((X_{n+1} = y)|X_n), \qquad \forall\, y \in E$$

ce qui signifie que la loi conditionnelle de X_{n+1} connaissant l'histoire du processus jusqu'à l'instant n (c'est à dire $\mathcal{F}_n$) est en fait déterminée par X_n c'est à dire l'état à l'instant n. C'est pourquoi ces processus sont quelquefois appelés "processus sans mémoire".

En utilisant la caractérisation de l'espérance conditionnelle relativement à une tribu engendrée par un système de variables aléatoires (voir cor. 3.2.8) on obtient:

Proposition 4.3.2. *Soit Q une matrice de transition sur E. Une suite X_n définie sur $(\Omega, \mathcal{F}, \mathbb{P})$, et à valeurs dans E est une chaîne de Markov (homogène) de matrice de transition Q si, pour tout système $f_k \in \mathcal{E}_b$, $k = 0, 1, \ldots, n+1$ et tout $n \geq 0$ on a:*

$$\mathbb{E}(\prod_{k=0}^{n+1} f_k(X_k)) = \mathbb{E}(Qf_{n+1}(X_n)\prod_{k=0}^{n} f_k(X_k)) \tag{4.4}$$

Il est aussi indiqué dans le cor. 3.2.8, qu'il suffit de prendre des fonctions f_k de la forme indicatrice d'ensembles dans $\mathcal{E}$. Mais, E étant dénombrable, on peut se contenter d'indicatrices de points, et ceci conduit à la définition plus élémentaire suivante.

Définition 4.3.3. *Soit Q une matrice de transition sur E. Une suite X_n définie sur $(\Omega, \mathcal{F}, \mathbb{P})$, et à valeurs dans E est une chaîne de Markov (homogène) de matrice de transition Q si, pour tous $x_0, x_1, \ldots, x_{n+1} \in E$ et tout $n \geq 0$,*

$$\mathbb{P}(X_0 = x_0, \ldots, X_{n+1} = x_{n+1}) = \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n)Q(x_n, x_{n+1}). \tag{4.5}$$

En résumé, si X est une chaîne de Markov par rapport aux $\mathcal{F}_n$ au sens de la def.4.3.1, X est une chaîne de Markov au sens de la def.4.3.3 et si X est une chaîne de Markov au sens de la def.4.3.3, X est une chaîne de Markov par rapport aux $\mathcal{F}_n^0$ au sens de la def.4.3.1. Les points de E s'appellent les états de la chaîne.

En fait la loi de la chaîne de Markov X est entièrement déterminée par sa loi initiale et sa matrice de transition comme le montre le théorème suivant qui est une simple conséquence de la discussion ci-dessus.

Théorème 4.3.4. *Soit $X = (\Omega, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P})$ un processus à valeurs dans E.*

- *Si X est une chaîne de Markov de loi initiale μ et de matrice de transition Q, on a, pour tous $x_0, \dots, x_n \in E$,*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n). \tag{4.6}$$

- *Réciproquement si le processus X vérifie (4.6) pour μ probabilité et Q matrice de transition sur E, alors X est une chaîne de Markov de loi initiale μ et de matrice de transition Q.*

La formule (4.6) a d'importantes conséquences. On a par exemple

$$\mathbb{P}(X_0 = x, X_2 = y) = \mu(x) \sum_z Q(x,z)Q(z,y) = \mu(x)Q^2(x,y)$$

et, de même, on montre par récurrence que

$$\mathbb{P}(X_0 = x, X_n = y) = \mu(x)Q^n(x,y). \tag{4.7}$$

En particulier

$$\mathbb{P}(X_n = y) = \sum_x \mu(x)Q^n(x,y) = \mu Q^n(y), \tag{4.8}$$

ce qui montre que, si X_0 a pour loi μ, X_n a pour loi μQ^n. On déduit également de (4.7) que, si $\mu(x) > 0$,

$$\mathbb{P}(X_n = y \,|\, X_0 = x) = \frac{\mathbb{P}(X_0 = x, X_n = y)}{\mathbb{P}(X_0 = x)} = Q^n(x,y). \tag{4.9}$$

Par ailleurs, si f est une application de E^{n+1} dans $\mathbb{R}$, positive ou bornée, on a (il suffit de sommer dans (4.6))

$$\mathbb{E}[f(X_0, \dots, X_n)] = \sum_{x_0, \dots, x_n} f(x_0, \dots, x_n)\mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n). \tag{4.10}$$

Si, dans (4.10), on choisit $f(x_0, \dots, x_n) = \mathbb{1}_{\{A_0\}}(x_0) \dots \mathbb{1}_{\{A_n\}}(x_n)$, $A_i \subset E$, on obtient,

$$\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n) = \sum_{x_0 \in A_0, \dots, x_n \in A_n} \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n). \tag{4.11}$$

Si, dans (4.11), on choisit $A_i = \{x_i\}$ pour certains indices et $A_j = E$ pour d'autres, on a, pour $l_0, \dots, l_k \in \mathbb{N}$,

$$\mathbb{P}(X_{l_0} = x_0, X_{l_0+l_1} = x_1, \dots, X_{l_0+l_1+\dots+l_k} = x_k) = \mu Q^{l_0}(x_0)Q^{l_1}(x_0, x_1) \dots Q^{l_k}(x_{k-1}, x_k). \tag{4.12}$$

L'exemple générique de chaînes de Markov

Théorème 4.3.5. *Soit $U_1, \ldots, U_n, \ldots$ une suite de v.a. à valeurs $(G, \mathcal{G})$ indépendantes et de même loi, X_0 une v.a. à valeurs E indépendante de la suite U_n et g une application de $E \times G$ dans E telle que, pour tout $x \in E$, $u \mapsto g(x, u)$ soit $\mathcal{G}$-mesurable. On pose, pour tout $n \geq 0$,*

$$X_{n+1} = g(X_n, U_{n+1}). \tag{4.13}$$

Alors X_n est une chaîne de Markov de matrice de transition

$$Q(x, y) = \mathbb{P}(g(x, U_1) = y).$$

`Preuve:`
Soit $\mathcal{F}_n = \sigma(X_0, U_1, \ldots, U_n)$ et $f \in \mathcal{E}^+$. On a d'après la prop.3.4.6

$$\mathbb{E}(f(X_{n+1}) \,|\, \mathcal{F}_n) = \mathbb{E}(f(g(X_n, U_{n+1})) \,|\, \mathcal{F}_n) = \phi(X_n)$$

où $\phi(x) = \mathbb{E}(f(g(x, U_{n+1}))) = \mathbb{E}(f(g(x, U_1)) = Qf(x)$ ∎

Un cas particulier important est le suivant. On considère une suite $U_1, \ldots, U_n, \ldots$ de v.a. à valeurs $\mathbb{Z}^d$ indépendantes et de même loi ν et indépendantes d'une variable aléatoire X_0. Pour $n \geq 0$, on pose

$$X_{n+1} = X_n + U_{n+1}, \tag{4.14}$$

$X_n = X_0 + U_1 + \ldots + U_n$ est donc une chaîne de Markov de matrice de transition $Q(x, y) = \mathbb{P}(x + U_1 = y) = \nu(y - x)$. Ce processus s'appelle la marche aléatoire de loi ν. On peut remarquer que dans ce cas on a $Q^n(x, y) = \nu^{(*n)}(y - x)$, où $\nu^{(*n)}$ est la $n$ième puissance de convolution de ν c'est à dire la loi de $U_1 + \ldots, +U_n$. Il est aussi facile de constater que si une matrice de transition Q sur $\mathbb{Z}^d$ est invariante par translation, c'est à dire $Q(x + z, y + z) = Q(x, y)$ pour tous points x,y,z alors c'est la matrice de transition d'une marche aléatoire de loi $\nu(y) = Q(0, y)$.

Chaînes de Markov stationnaires et réversibles

Définition 4.3.6. *Une chaîne X_n est dite stationnaire si, pour tous $k, n \in \mathbb{N}$, la loi de $(X_0, \ldots, X_n)$ est égale à la loi de $(X_k, \ldots, X_{n+k})$. Elle est dite réversible si, pour tout entier $T \geq 1$, la loi du vecteur $(X_0, X_1, \ldots, X_T)$ est identique à la loi du vecteur "retourné" $(X_T, X_{T-1}, \ldots, X_0)$.*

Définition 4.3.7. *On dit qu'une mesure positive λ sur E est excessive (ou surharmonique) si $\lambda \geq \lambda Q$, invariante (ou harmonique) si $\lambda = \lambda Q$. On dit que λ est réversible si pour tous $x, y \in E$, $\lambda(x)Q(x, y) = \lambda(y)Q(y, x)$.*

On voit immédiatement qu'une mesure réversible est invariante car si λ est réversible

$$\lambda Q(y) = \sum_x \lambda(x)Q(x, y) = \sum_x \lambda(y)Q(y, x) = \lambda(y) \sum_x Q(y, x) = \lambda(y)$$

Il est en général assez facile de calculer une mesure réversible. Cependant elle n'existe pas toujours même s'il existe une mesure invariante. On verra qu'une probabilité invariante est un objet particulièrement intéressant. Une des raisons est:

Proposition 4.3.8. *Si λ est une probabilité invariante, la chaîne X_n de loi initiale λ est stationnaire.*

`Preuve:`
Soient $x_0, \dots, x_n \in E$. On a, d'après 4.12, pour tout k,

$$\begin{aligned}\mathbb{P}(X_k = x_0, \dots, X_{n+k} = x_n) &= \lambda Q^k(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)\\ &= \lambda(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)\\ &= \mathbb{P}(X_0 = x_0, \dots, X_n = x_n)\end{aligned}$$

Proposition 4.3.9. *Si λ est une probabilité réversible, la chaîne X_n de loi initiale λ est réversible.*

`Preuve:`
Soient $x_0, \dots, x_T \in E$. On a,

$$\begin{aligned}\mathbb{P}(X_0 = x_0, \dots, X_T = x_T) &= \lambda(x_0)Q(x_0, x_1)\dots Q(x_{T-1}, x_T)\\ &= Q(x_1, x_0)\lambda(x_1)Q(x_1, x_2)\dots Q(x_{T-1}, x_T)\\ &= \dots = Q(x_1, x_0)Q(x_2, x_1)\dots Q(x_T, x_{T-1})\lambda(x_T)\\ &= \mathbb{P}(X_0 = x_T, \dots, X_T = x_0)\end{aligned}$$

4.4. Chaînes canoniques

Intuitivement étant données une probabilité μ et une matrice de transition Q sur E, il existe une chaîne de Markov de loi initiale μ et de matrice de transition Q. En effet on tire X_0 selon la loi μ, on obtient x_0. Alors on tire X_1 selon la loi $Q(x_0, .)$, on obtient x_1, puis on tire X_2 selon la loi $Q(x_1, .)$, on obtient x_2 et ainsi de suite. C'est ce que confirme le théorème suivant.

Théorème 4.4.1. *Soient μ une probabilité et Q une matrice de transition sur E. Considérons l'espace canonique*

$$\Omega = E^{\mathbb{N}},\ \omega = (\omega_n)_{n\geq 0},\ X_n(\omega) = \omega_n,\ \mathcal{F}_n = \sigma(X_k, k \leq n),\ \mathcal{F} = \sigma(X_k, k \geq 0). \quad (4.15)$$

Alors il existe une unique probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ telle que $(\Omega, \mathcal{F}_n, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P})$ soit une chaîne de Markov de loi initiale μ et de matrice de transition Q.

`Preuve:`
On définit une probabilité μ_n sur E^{n+1} par

$$\mu_n(x_0, \dots, x_n) = \mu(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n).$$

Ces probabilités vérifient la condition (4.1) du th.4.1.4. Donc il existe une unique probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ telle que

$$\mathbb{P}(X_0 = x_0, \dots, x_n) = \mu_n(x_0, \dots, x_n) = \mu(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)$$

On conclut par le th.4.3.4 ∎

On note $\mathbb{P}_\mu$ la probabilité relative à loi initiale μ. Si $\mu = \delta_x$, δ_x étant la mesure de Dirac du point, on note $\mathbb{P}_x$ pour $\mathbb{P}_{\delta_x}$. En vertu de (4.6), on a pour tout $A \in \mathcal{F}$,

$$\mathbb{P}_\mu(A) = \sum_{x\in E} \mu(x)\mathbb{P}_x(A). \tag{4.16}$$

On peut maintenant introduire l'objet sur lequel nous allons dorénavant travailler.

Définition 4.4.2. *Soit Q une matrice de transition sur E. On appelle chaîne de Markov canonique de matrice de transition Q, un processus $(\Omega, \mathcal{F}_n, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P}_\mu)$ sur l'espace canonique défini par (4.15) et où, pour chaque $\mu \in \mathcal{M}_1$, $\mathbb{P}_\mu$ est une probabilité sur $(\Omega, \mathcal{F})$ vérifiant, pour tous $n, x_0, \dots, x_n$,*

$$\mathbb{P}_\mu(X_0 = x_0 \dots, X_n = x_n) = \mu x_0 Q(x_0, x_1) \dots Q(x_{n-1}, x_n). \tag{4.17}$$

Alors, pour chaque $\mu \in \mathcal{M}_1$, $(\Omega, \mathcal{F}_n, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P}_\mu)$ est une chaîne de Markov de loi initiale μ et de matrice de transition Q. Dans ce cadre, on a les formules suivantes vu (4.9). Pour tous $n \geq 0$, $g \in \mathcal{E}^+$, $\mu \in \mathcal{M}_1$,

$$\mathbb{P}_x(X_n = y) = Q^n(x, y), \quad \mathbb{E}_x(g(X_n)) = Q^n g(x) \quad \mathbb{P}_\mu(X_n = y) = \mu Q^n(y), \tag{4.18}$$

$$\mathbb{E}_\mu(g(X_n)) = \sum_x \mu(x) Q^n g(x) = \mu Q^n g \tag{4.19}$$

Evidemment une chaîne de Markov n'est pas toujours donnée sous forme canonique. Cependant si $X' = (\Omega', \mathcal{F}'_n, \mathcal{F}', (X'_n)_{n\geq 0}, \mathbb{P}')$ une chaîne de Markov de loi initiale μ et de matrice de transition Q et si $X = (\Omega, \mathcal{F}_n, \mathcal{F}, (X_n)_{n\geq 0}, \mathbb{P}_\mu)$ est la chaîne canonique de loi initiale μ et de matrice de transition Q, les processus X et X' ont même loi (voir 4.1.1.2) et on peut effectuer tous les calculs sur la chaîne X. Par exemple, on a $\mathbb{P}'(X'_n \text{ visite } A) = \mathbb{P}_\mu(X_n \text{ visite } A)$.

La propriété de Markov

Le principal intérêt de se placer sur l'espace canonique est l'existence d'un opérateur de translation. On considère l'espace canonique (4.15). On définit une application θ de Ω dans Ω par $\omega = (\omega_n)_{n\geq 0} \mapsto \theta(\omega) = (\omega_{n+1})_{n\geq 0}$.
Cette application s'appelle l'opérateur de translation sur Ω. On pose alors

$$\theta_0 = \text{Id.}, \ \ \theta_1 = \theta, \ \ \theta_n = \theta_{n-1} \circ \theta = \theta \circ \theta_{n-1}.$$

On a donc $X_k \circ \theta_n = X_{n+k}$ pour tout $k \geq 0$. On a en particulier, pour tous p et n,

$$\theta_n^{-1}(\{X_0 = x_0, \dots, X_p = x_p\}) = \{X_n = x_0, \dots, X_{n+p} = x_p\}$$

ce qui montre que θ_n est une application mesurable de $(\Omega, \sigma(X_k, k \geq n))$ dans $(\Omega, \mathcal{F})$ et que, pour tous n et p,

$$\theta_n^{-1}(\mathcal{F}_p) = \sigma(X_n, \dots, X_{n+p}). \tag{4.20}$$

On peut alors énoncer:

Théorème 4.4.3. *Soit X une chaîne de Markov canonique. On a, pour tout $n \geq 0$, toute loi initiale μ et toute $\Psi \in b\mathcal{F}$ (resp. $\mathcal{F}^+$)*

$$\boxed{\mathbb{E}_\mu(\Psi \circ \theta_n \,|\, \mathcal{F}_n) = \mathbb{E}_{X_n}(\Psi) \ , \ \mathbb{P}_\mu \ p.s.} \tag{4.21}$$

ce qui équivaut à, pour toute $\Phi \in b\mathcal{F}_n$, toute $\Psi \in b\mathcal{F}$, (resp. toute $\Phi \in \mathcal{F}_n^+$, toute $\Psi \in \mathcal{F}^+$)

$$\mathbb{E}_\mu[\Phi\Psi \circ \theta_n] = \mathbb{E}_\mu\left(\Phi\mathbb{E}_{X_n}(\Psi)\right)$$

`Preuve:`
La relation (4.21) est vraie pour $\Phi = \mathbb{1}_{\{X_0=a_0,\dots,X_n=a_n\}}$ et $\Psi = \mathbb{1}_{\{X_0=b_0,\dots,X_k=b_k\}}$, les deux termes étant alors égaux à

$$\mu(a_0)Q(a_0, a_1)\dots Q(a_{n-1}, a_n)\mathbb{1}_{\{a_n\}}(b_0)Q(b_0, b_1)\dots Q(b_{k-1}, b_k).$$

Par sommation, on en déduit (4.21) pour $\Phi \in \mathcal{F}_n^+$ et $\Psi \in \mathcal{F}_k^+$, k quelconque. On conclut facilement en utilisant le cor.3.2.8 ∎

Ce résultat s'étend sans peine aux temps d'arrêt. Considérons maintenant un temps d'arrêt T. Evidemment par temps d'arrêt, on entend temps d'arrêt de la filtration $\mathcal{F}_n$ de l'espace canonique (4.15). On définit θ_T de $\{T < +\infty\}$ dans Ω par la formule $\theta_T(\omega) = \theta_{T(\omega)}(\omega)$ On a alors sur l'ensemble $\{T < +\infty\}$, et pour tout $n \geq 0$,

$$X_n \circ \theta_T = X_{n+T}, \qquad \theta_T^{-1}(\{X_0 = x_0, \dots, X_n = x_n\}) = \{X_T = x_0, \dots, X_{T+n} = x_n\}.$$

Théorème 4.4.4. *Soient X une chaîne de Markov canonique et T un temps d'arrêt. On a, pour toute loi initiale μ et toute $\Psi \in b\mathcal{F}$ (resp. $\Psi \in \mathcal{F}^+$)*

$$\boxed{\mathbb{E}_\mu(\mathbb{1}_{\{T<+\infty\}}\Psi \circ \theta_T \,|\, \mathcal{F}_T) = \mathbb{1}_{\{T<+\infty\}}\mathbb{E}_{X_T}(\Psi) \ , \ \mathbb{P}_\mu \ p.s.} \tag{4.22}$$

ce qui équivaut à, pour toute $\Phi \in b\mathcal{F}_T$, toute $\Psi \in b\mathcal{F}$, (resp. toute $\Phi \in \mathcal{F}_T^+$, toute $\Psi \in \mathcal{F}^+$)

$$\mathbb{E}_\mu[\mathbb{1}_{\{T<+\infty\}}\Phi\Psi \circ \theta_T] = \mathbb{E}_\mu[\mathbb{1}_{\{T<+\infty\}}\Phi\mathbb{E}_{X_T}(\Psi)]$$

`Preuve:`
Soient $\Phi \in \mathcal{F}_T^+$ et $\Psi \in \mathcal{F}^+$. On remarque que $\Phi\mathbb{1}_{\{T=n\}} \in \mathcal{F}_n^+$. On a alors

$$\mathbb{E}_\mu(\mathbb{1}_{\{T<+\infty\}}\Phi\Psi \circ \theta_T) = \sum_{n\geq 0} \mathbb{E}_\mu(\mathbb{1}_{\{T=n\}}\Phi\Psi \circ \theta_n)$$

$$= \sum_{n\geq 0} \mathbb{E}_\mu(\mathbb{1}_{\{T=n\}}\Phi\mathbb{E}_{X_n}(\psi)) = \mathbb{E}_\mu[\mathbb{1}_{\{T<+\infty\}}\Phi\mathbb{E}_{X_T}(\Psi)] \ ∎$$

Evidemment le th.4.4.4 implique le th.4.4.3. Traditionnellement la propriété (4.21) s'appelle la propriété de Markov et la propriété (4.22) s'appelle la propriété forte de Markov. Vu que $\mathbb{1}_{\{\theta_T^{-1}(B)\}} = \mathbb{1}_{\{B\}} \circ \theta_T$, (4.22) s'écrit pour des ensembles:

$$\mathbb{P}_\mu[\{T < +\infty\} \cap A \cap \theta_T^{-1}(B)] = \mathbb{E}_\mu[\mathbb{1}_{\{T<+\infty\}}\mathbb{1}_{\{A\}}\mathbb{P}_{X_T}(B)],\ A \in \mathcal{F}_T,\ B \in \mathcal{F}. \quad (4.23)$$

Simulation d'une chaîne de Markov donnée par sa fonction de transition

On présente une deuxième méthode pour construire une chaîne de Markov de loi initiale μ et de matrice de transition Q données. Elle a le double intérêt de montrer que toute chaîne est de la forme (4.13) et de fournir un moyen de simuler une chaîne à partir d'une suite de nombres aléatoires compris entre 0 et 1 (surtout si E est fini). On suppose $E = \mathbb{N}$, cas auquel on peut toujours se ramener en numérotant les points de E. On pose, pour tous $i \in \mathbb{N}$ et $x \in [0,1]$,

$$s_0 = 0, s_1 = \mu(0), \ldots, s_n = \sum_{k=0}^{n-1} \mu(k), \ldots$$

$$s_0(i) = 0, s_1(i) = Q(i,0), \ldots, s_n(i) = \sum_{k=0}^{n-1} Q(i,k), \ldots$$

$$g(x) = k \text{ si } x \in [s_k, s_{k+1}[,\ g(1) = 0,$$

$$g(i,x) = k \text{ si } x \in [s_k(i), s_{k+1}(i)[,\ g(i,1) = 0.$$

Evidemment, si $\mu(k) = 0$, on a $s_k = s_{k+1}$ auquel cas $[s_k, s_{k+1}[= \emptyset$. De même pour les $s_k(i)$.

Soit $U_0, U_1, \ldots, U_n, \ldots$ une suite de v.a. indépendantes de même loi uniforme sur $[0,1]$. On pose

$$X_0 = g(U_0), \quad X_{n+1} = g(X_n, U_{n+1}).$$

On sait (th.4.3.5) que X_n est une chaîne de Markov de matrice de transition $\mathbb{P}(g(i,U_1) = j)$. Mais on a

$$\mathbb{P}(X_0 = j) = \mathbb{P}(g(U_0) = j) = \mathbb{P}(U_0 \in [s_j, s_{j+1}[) = s_{j+1} - s_j = \mu(j),$$

$$\mathbb{P}(g(i,U_1) = j) = \mathbb{P}(U_1 \in [s_j(i), s_{j+1}(i)[) = s_{j+1}(i) - s_j(i) = Q(i,j).$$

Donc X_n est une chaîne de Markov de loi initiale μ et de matrice de transition Q.

4.5. Récurrence et transience

On reprend, dans le cadre d'une chaîne canonique, les définitions de temps d'entrée, temps de retour successifs, nombre de passages d'un processus dans un ensemble données dans 4.1.8. L'utilisation de l'opérateur de translation θ permet d'obtenir un certain nombre de relations évidentes:

- $S_A = 1 + T_A \circ \theta$, $\mathbb{P}_x(T_A = 0) = 1$ pour $x \in A$, $\mathbb{P}_x(T_A = S_A) = 1$ si $x \notin A$.

- Sur l'ensemble $\{S_A^n < \infty\}$ on a $S_A^{n+1} = S_A^n + S_A \circ \theta_{S_A^n} = S_A + S_A^n \circ \theta_{S_A}$.

- Dans le cas où A est réduit à un point $x \in E$, on a:

$$N_x = 1 + \sum_{n=1}^{\infty} \mathbb{1}_{\{S_x^n < +\infty\}}, \ \mathbb{P}_x \ p.s. \tag{4.24}$$

Définition 4.5.1. *On appelle matrice potentielle de Q (ou de X) la matrice*

$$U = I + Q + \ldots + Q^n + \ldots = \sum_{k=0}^{+\infty} Q^k$$

La matrice potentielle peut s'interpréter à l'aide la chaîne X car

$$\sum_{k=0}^{\infty} Q^k(x,y) = \sum_{k=0}^{\infty} \mathbb{P}_x(X_k = y) = \sum_{k=0}^{\infty} \mathbb{E}_x(\mathbb{1}_{\{y\}} \circ X_k) = \mathbb{E}_x(N_y).$$

Donc $U(x,y)$ est le nombre moyen de visites à y de la chaîne partant de x. Les termes $U(x,y)$ et $\mathbb{P}_x(T_y < +\infty)$ sont liés par:

Proposition 4.5.2. *Pour tout couple x, y on a*

$$U(x,y) = \mathbb{P}_x(T_y < +\infty)\, U(y,y).$$

`Preuve:`
On a en appliquant la propriété forte de Markov (4.22),

$$\begin{aligned}
U(x,y) &= \mathbb{E}_x(N_y) = \mathbb{E}_x(N_y \mathbb{1}_{\{T_y<\infty\}}) = \mathbb{E}_x\Big(\sum_{n \geq T_y} (\mathbb{1}_{\{y\}} \circ X_n) \mathbb{1}_{\{T_y<\infty\}}\Big) \\
&= \sum_{n \geq 0} \mathbb{E}_x(\mathbb{1}_{\{T_y<\infty\}} (\mathbb{1}_{\{y\}} \circ X_n) \circ \theta_{T_y}) \\
&= \sum_{n \geq 0} \mathbb{E}_x(\mathbb{1}_{\{T_y<\infty\}} \mathbb{E}_{X_{T_y}}(\mathbb{1}_{\{y\}} \circ X_n)) = \sum_{n \geq 0} \mathbb{E}_x(\mathbb{1}_{\{T_y<\infty\}} \mathbb{E}_y(\mathbb{1}_{\{y\}} \circ X_n)) \\
&= \mathbb{P}_x(T_y < \infty) \sum_{n \geq 0} \mathbb{E}_y(\mathbb{1}_{\{y\}} \circ X_n) = \mathbb{P}_x(T_y < \infty) U(y,y) \ \blacksquare
\end{aligned}$$

Remarque Cette proposition montre que la fonction $x \mapsto U(x,y)$ atteint un maximum au point y. Cette propriété est parfois appelée "principe du maximum".

Définition 4.5.3. *On pose $a(x) = \mathbb{P}_x(S_x < \infty)$. On dit que x est récurrent si $a(x) = 1$ et transient si $a(x) < 1$.*

Le résultat fondamental est:

Théorème 4.5.4. *Soit $x \in E$.*

1. *Si x est récurrent alors $\mathbb{P}_x(N_x = +\infty) = 1$ et $U(x,x) = +\infty$.*
2. *Si x est transient alors $\mathbb{P}_x(N_x < +\infty) = 1$ et $U(x,x) = (1-a(x))^{-1} < \infty$. De plus, sous $\mathbb{P}_x$, la variable aléatoire N_x, à valeurs dans $\mathbb{N}^*$, suit une loi géométrique de paramètre $a(x)$.*

`Preuve:`
On a, utilisant (4.23)

$$\begin{aligned}\mathbb{P}_x(S_x^n < +\infty) &= \mathbb{P}_x(S_x^{n-1} < +\infty,\ S_x \circ \theta_{S_x^{n-1}} < +\infty) \\ &= \mathbb{P}_x(\{S_x^{n-1} < +\infty\} \cap \theta^{-1}_{S_x^{n-1}}(\{S_x < +\infty\})) \\ &= \mathbb{E}_x(\mathbb{1}_{\{S_x^{n-1}<+\infty\}} \mathbb{P}_{X_{S_x^{n-1}}}(S_x < +\infty)) \\ &= \mathbb{P}_x(S_x^{n-1} < +\infty)\mathbb{P}_x(S_x < +\infty) = (a(x))^n\end{aligned}$$

• 1. Si $a(x) = 1$, alors, pour tout n, $\mathbb{P}_x(S_x^n < +\infty) = 1$. Vu (4.24), on a $\mathbb{P}_x(N_x = +\infty) = 1$ et a fortiori $U(x,x) = \mathbb{E}_x(N_x) = +\infty$.

• 2. Si $a(x) < 1$, alors, vu (4.24),

$$U(x,x) = \mathbb{E}_x(N_x) = 1 + \sum_{n=1}^{\infty} \mathbb{P}_x(S_x^n < +\infty) = (1-a(x))^{-1} < +\infty$$

et a fortiori $\mathbb{P}_x(N_x < +\infty) = 1$. De plus on a $\mathbb{P}_x(N_x > n) = \mathbb{P}_x(S_x^n < +\infty) = (a(x))^n$ et donc N_x suit une loi géométrique de paramètre $a(x)$ sous $\mathbb{P}_x$. ∎

Autrement dit, partant d'un point x, l'on revient une infinité de fois en ce point avec une probabilité égale à 0 ou 1. On retrouve donc une certaine forme de la loi du 0 ou 1 puisque l'événement $(N_x = \infty)$ est mesurable par rapport à toutes les tribus $\mathcal{B}_n = \sigma(X_n, X_{n+1}, \ldots)$ et donc par rapport à leur intersection. Ceci constitue l'un des résultats les plus remarquables sur le comportement des chaînes de Markov.

Proposition 4.5.5. *Si E est fini, il existe au moins un état récurrent.*

`Preuve:`
En effet $\sum_{y\in E} N_y = N_E = +\infty$ d'où $\sum_{y\in E} U(x,y) = \mathbb{E}_x(\sum_{y\in E} N_y) = +\infty$. Il existe donc $y \in E$ tel que $U(x,y) = +\infty$ mais (prop.4.5.2) $U(y,y) \geq U(x,y) = +\infty$ et y est récurrent ∎

Supposons que la chaîne parte de x. Alors son comportement est régi par la loi $\mathbb{P}_x$. Elle revient en x à l'instant S_x. D'après la propriété forte de Markov, son comportement après S_x est régi par la loi $\mathbb{P}_{X_{S_x}}$ mais $\mathbb{P}_{X_{S_x}} = \mathbb{P}_x$ et donc son comportement après S_x est indépendant de son comportement avant S_x et est régi par la même loi. C'est ce que formalise la proposition suivante.

Proposition 4.5.6. *Soit x un point récurrent et $Y_0, Y_1, \ldots, Y_n$ des v.a.r. $\mathcal{F}_{S_x}$-mesurables. Pour $0 \leq k \leq n$ on note ρ_k la loi de Y_k sous $\mathbb{P}_x$. Alors, sous $\mathbb{P}_x$, les v.a. $Y_0, Y_1 \circ \theta_{S_x^1}, \ldots, Y_n \circ \theta_{S_x^n}$ (qui sont bien définies $\mathbb{P}_x$ p.s.) sont indépendantes et de lois respectives $\rho_0, \rho_1, \ldots, \rho_n$.*

`Preuve:`
Vu que $\theta_{S_x^n} = \theta_{S_x^{n-1}} \circ \theta_{S_x^1}$, on a pour des fonctions positives $f_0, \dots, f_n$:

$$\begin{aligned}
\mathbb{E}_x(\prod_{k=0}^n f_k(Y_k \circ \theta_{S_x^k})) &= \mathbb{E}_x\left(f_0(Y_0)\prod_{k=1}^n (f_k(Y_k \circ \theta_{S_x^{k-1}})) \circ \theta_{S_x^1}\right) \\
&= \mathbb{E}_x(f_0(Y_0))\mathbb{E}_x\left(\prod_{k=1}^n f_k(Y_k \circ \theta_{S_x^{k-1}})\right) \\
&= \mathbb{E}_x(f_0(Y_0))\mathbb{E}_x(f_1(Y_1))\mathbb{E}_x\left(\prod_{k=2}^n f_k(Y_k \circ \theta_{S_x^{k-2}})\right) \\
&= \ldots = \prod_{k=0}^n \mathbb{E}_x(f_k(Y_k)) = \prod_{k=0}^n \rho_k(f_k)
\end{aligned}$$

d'où le résultat annoncé ∎

Remarque 4.5.7. *Cette proposition montre que les tribus $\mathcal{F}_{S_x}, \theta_{S_x}^{-1}(\mathcal{F}_{S_x}), \dots, \theta_{S_x^n}^{-1}(\mathcal{F}_{S_x})$ sont indépendantes sous $\mathbb{P}_x$.*

Cette proposition sera surtout utilisée dans le cas particulier suivant:

Corollaire 4.5.8. *Soit f une fonction sur E et x un point récurrent. On pose $Z_0 = \sum_{k=0}^{S_x-1} f(X_k)$ puis $Z_n = Z_0 \circ \theta_{S_x^n}$. Alors $Z_n = \sum_{k=S_x^n}^{S_x^{n+1}-1} f(X_k)$ et, sous $\mathbb{P}_x$, les variables aléatoires Z_n, $n \geq 0$ sont i.i.d.*

En fait, le même argument montre que les excursions successives, c'est à dire les portions de trajectoires $E_n = (X_{S_x^n}, \dots, X_{S_x^{n+1}-1})$, séparant deux passages au point x, sont, sous $\mathbb{P}_x$, des variables aléatoires indépendantes et de même loi si on les considère comme des éléments à valeurs dans l'espace dénombrable des mots finis, non vides, construits sur l'alphabet E. C'est cette propriété fondamentale qui est à la base d'une bonne part de la théorie des processus de Markov.

Nous allons maintenant passer aux problèmes de communication entre points de E.

Irréductibilité

L'étude suivante va permettre de décomposer l'espace des états E en classes de points de même nature. En fait on se restreindra souvent, dans la suite, au cas où E sera formé d'une seule classe, que l'on appellera "cas irréductible".

Définition 4.5.9. *On dit que x conduit à y, ce qu'on note $x \to y$, si $\mathbb{P}_x(T_y < +\infty) > 0$. On dit que la chaîne est irréductible si $x \to y$ pour tout couple (x, y).*

Vu que $\{T_y < +\infty\} = \cup_{n\geq 0}\{X_n = y\}$ on obtient immédiatement:

Lemme 4.5.10. *$x \to y$ est équivalent à l'une des conditions suivantes:*

1. *Il existe $n \geq 0$ tel que $Q^n(x, y) > 0$*

2. *Il existe une suite* $x_0 = x, x_1, \ldots, x_{n-1}, x_n = y$ *avec* $Q(x_{k-1}, x_k) > 0$ *pour* $k \geq 1$

3. $U(x, y) > 0$

Lemme 4.5.11. *Soit x,y,z trois points de E alors:*

$$\mathbb{P}_x(T_z < +\infty) \geq \mathbb{P}_x(T_y < +\infty)\mathbb{P}_y(T_z < +\infty)$$

En conséquence si $x \to y$ *et si* $y \to z$*, alors* $x \to z$*. La relation binaire* $x\mathcal{R}y$*, si* $x \to y$ *et* $y \to x$ *est donc une relation d'équivalence sur* E*.*

`Preuve:`
On a, utilisant (4.23),

$$\mathbb{P}_x(T_z < +\infty) \geq \mathbb{P}_x(T_y + T_z \circ \theta_{T_y} < +\infty)$$
$$= \mathbb{P}_x(\{T_y < +\infty\} \cap \theta_{T_y}^{-1}(T_z < +\infty)) = \mathbb{P}_x(T_y < +\infty)\mathbb{P}_y(T_z < +\infty) > 0 \ \blacksquare$$

Nous allons maintenant préciser cette décomposition en classes d'équivalence.

Proposition 4.5.12. *Soit x un point récurrent et y tel que* $x \to y$*. Alors:*

1. *Le point y est récurrent et* $y \to x$*.*

2. *On a* $(\mathbb{P}_x(N_y) = \infty) = (\mathbb{P}_y(N_x) = \infty) = 1$

`Preuve:`
On peut supposer $x \neq y$ car sinon il n'y a rien à prouver.
On reprend la construction du Corollaire 4.5.8 avec $Z_0 = \sum_{k=0}^{S_x-1} \mathbb{1}_{\{y\}} \circ X_k$. L'événement $(Z_n > 0)$ est égal à $\{\exists k\ ;\ S_x^n \leq k < S_x^{n+1}\ ;\ X_k = y\}$. Ces événements sont indépendants et de même probabilité soit α sous $\mathbb{P}_x$. On ne peut avoir $\alpha = 0$ car alors $\mathbb{P}_x(T_y < \infty) = \mathbb{P}_x(\cup_n(Z_n > 0))$ serait nul. Il résulte alors du Lemme de Borel Cantelli que $\mathbb{P}_x$ p.s., l'événement $(Z_n > 0)$ se produit une infinité de fois et donc $\mathbb{P}_x(N_y = \infty) = 1$ d'où $U(x, y) = \mathbb{E}_x(N_y) = \infty$. De la relation $U(x, y) = \mathbb{P}_x(T_y < \infty)U(y, y)$ on déduit que $U(y, y) = \infty$ et donc que y est récurrent. Il reste donc seulement à prouver que $y \to x$. Pour ceci on remarque que

$$(T_y < \infty) \cap (T_x \circ \theta_{T_y} = \infty) \subset (N_x < \infty), \quad \mathbb{P}_x \text{ p.s.}$$

et par application de la propriété de Markov forte on obtient:

$$\mathbb{P}_x(T_y < \infty)\mathbb{P}_y(T_x = \infty) \leq \mathbb{P}_x(N_x < \infty) = 0$$

et donc $\mathbb{P}_y(T_x = \infty) = 0$ puisque $\mathbb{P}_x(T_y < \infty) > 0 \ \blacksquare$

Remarque 4.5.13. *La démonstration ci dessus, introduisant les excursions a le mérite de se transposer aux cas de processus plus généraux. Dans notre cas, on peut faire une preuve plus élémentaire. Par exemple pour montrer que y est récurrent on peut écrire:*

Puisque $x \to y$, il existe $n \geq 1$ tel que $Q^n(x,y) > 0$. D'après le principe du maximum, 4.5.2, on peut écrire:

$$\begin{aligned} U(y,y) &\geq U(x,y) = \sum_k Q^k(x,y) \geq \sum_k Q^{k+n}(x,y) \\ &\geq \sum_k Q^k(x,x)Q^n(x,y) = U(x,x)Q^n(x,y) = +\infty \end{aligned}$$

Corollaire 4.5.14. *L'espace E se décompose donc en classes transitoires et classes récurrentes telles que:*

1. *x et y appartiennent à deux classes distinctes si et seulement si $U(x,y) = 0$ ou $U(y,x) = 0$.*
2. *Dans une classe transitoire on a $0 < U(x,y) < \infty$ pour tout couple (x,y)*
3. *Dans une classe récurrente on a $U(x,y) = \infty$ pour tout couple (x,y). De plus une classe récurrente est close, c'est à dire qu'en désignant par C une telle classe, si $x \in C$ alors $\mathbb{P}_x(\exists n \geq 1\ ;\ X_n \notin C) = 0$.*

`Preuve:`
Au vu de la proposition 4.5.12 et du principe du maximum, il ne reste qu'à prouver l'assertion relative aux classes closes. Soit C une classe de récurrence, $x \in C$ et $y \notin C$ tels qu'il existe $n \geq 1$ avec $\mathbb{P}_x(X_n = y) > 0$. On aurait donc $x \to y$ et par conséquent $y \to x$ d'après 4.5.12 ce qui est impossible puisque x et y ne sont pas dans la même classe ∎

A l'issue de cette étude on peut donc préciser le comportement d'une chaîne de Markov. Partant d'un point récurrent celle ci restera dans sa classe de récurrence et visitera une infinité de fois tous les points de la classe. Partant d'un point d'une classe transitoire, le comportement est moins clair: si la classe transitoire, soit T, est finie, la chaîne sortira presque sûrement de cette classe puisque $\mathbb{E}_x(N_T) = \sum_{y\in T} U(x,y) < \infty$ et donc $\mathbb{P}_x(N_T < \infty) = 1$. Dans ce cas la chaîne passera soit dans une autre classe transitoire soit dans une classe de récurrence où elle restera "piégée". Si la classe transitoire est infinie, le même raisonnement que ci-dessus montre que la chaîne sortira presque sûrement de tout sous ensemble fini de T.

En définitive, pour les chaînes irréductibles, on a la dichotomie suivante:

Corollaire 4.5.15. *Soit X une chaîne de Markov irréductible, alors*

- *Soit pour tous $x, y \in E$, $U(x,y) < +\infty$, alors tous les états sont transients et la chaîne est dite transiente,*
- *soit pour tous $x, y \in E$, $U(x,y) = +\infty$, alors tous les états sont récurrents et la chaîne est dite récurrente.*

Le critère de récurrence suivant est très utilisé.

Proposition 4.5.16. *Soit X une chaîne de Markov admettant une probabilité invariante λ.*

1. *Tout état y tel que $\lambda(y) > 0$ est récurrent.*
2. *Si, de plus, X est irréductible, alors X est récurrente.*

`Preuve:`
Soit λ une probabilité invariante et $y \in E$

$$\begin{aligned} \lambda U(y) &= \sum_n \lambda Q^n(y) = \sum_n \lambda(y) \\ \lambda U(y) &= \sum_x \lambda(x) U(x,y) \leq U(y,y) \sum_x \lambda(x) = U(y,y) \end{aligned}$$

On en déduit que si $\lambda(y) > 0$ alors $\lambda U(y) = \infty$ et donc aussi $U(y,y)$ ∎

Il faut bien noter que le fait que λ soit de masse totale finie est essentiel. Il se peut très bien qu'une chaîne transitoire admette une mesure invariante (mais qui sera nécessairement de masse totale infinie).

Le théorème ergodique

L'indépendance des excursions d'une chaîne de Markov est l'argument essentiel du théorème ergodique que nous allons maintenant établir.
On appelle théorèmes ergodiques des théorèmes concernant le comportement, lorsque $n \to +\infty$, de

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \text{ ou de } \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)}$$

Définition 4.5.17. *Soit x un point récurrent, on définit une mesure λ_x sur E par*

$$\lambda_x(y) = \mathbb{E}_x\Big(\sum_{k=0}^{S_x - 1} \mathbb{1}_{\{y\}} \circ X_k\Big),\ y \in E, \quad \textit{ou encore} \quad \lambda_x(f) = \mathbb{E}_x\Big(\sum_{k=0}^{S_x-1} f(X_k)\Big),\ f \in \mathcal{E}^+$$

On a en particulier $\lambda_x(1) = \mathbb{E}_x(S_x)$ et $\lambda_x(x) = 1$.

Théorème 4.5.18. *Soit x un point récurrent et $f, g \in L^1(\lambda_x)$ avec $\lambda_x(g) \neq 0$, alors:*

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \to \frac{\lambda_x(f)}{\lambda_x(g)} \quad \mathbb{P}_x \ p.s.$$

`Preuve:`
Soit f une fonction positive, λ_x intégrable, on pose

$$Z_0 = \sum_{k=0}^{S_x-1} f(X_k),\ Z_1 = Z \circ \theta_{S_x^1}, \ldots, Z_n = Z \circ \theta_{S_x^n}$$

La v.a. Z_0 est $\mathcal{F}_{S_x}$-mesurable et $\mathbb{E}_x(Z_0) = \lambda_x(f)$. D'après la prop.4.5.6, les variables aléatoires Z_k, $k \geq 0$, sont, sous $\mathbb{P}_x$, indépendantes et de même loi. D'après la loi forte des grands nombres,

$$\frac{1}{n}(Z_0 + \ldots + Z_{n-1}) \to_n \lambda_x(f) \;\; \mathbb{P}_x \text{ p.s.}$$

Comme, pour tout p,

$$Z_p = Z \circ \theta_{S_x^p} = (\sum_{k=0}^{S_x-1} f(X_k)) \circ \theta_{S_x^p} = \sum_{k=S_x^p}^{S_x^{p+1}-1} f(X_k)$$

on a

$$\frac{1}{n} \sum_{0 \leq k < S_x^n} f(X_k) = \frac{1}{n}(Z_0 + \ldots + Z_{n-1}) \to_n \lambda_x(f) \;\; \mathbb{P}_x \text{ p.s.}$$

Cette convergence a donc lieu aussi, pour toute suite strictement croissante d'entiers qui tend vers $+\infty$. A l'entier $m \in \mathbb{N}$, on associe l'unique entier $\nu(m)$ tel que $S_x^{\nu(m)} < m \leq S_x^{\nu(m)+1}$, alors $\nu(m) \to_m +\infty$ et $\sum_{0 \leq k < m} \mathbb{1}_{\{x\}}(X_k) = \nu(m) + 1$, $\mathbb{P}_x$ p.s., d'o ù:

$$\frac{\nu(m)}{\nu(m)+1} \frac{\sum_{0 \leq k < S_x^{\nu(m)}} f(X_k)}{\nu(m)} \leq \frac{\sum_{0 \leq k < m} f(X_k)}{\sum_{0 \leq k < m} \mathbb{1}_{\{x\}}(X_k)} \leq \frac{\sum_{0 \leq k < S_x^{\nu(m)+1}} f(X_k)}{\nu(m)+1}$$

les termes latéraux tendant vers $\lambda_x(f)$ $\mathbb{P}_x$ p.s., on a donc

$$\frac{\sum_{0 \leq k < m} f(X_k)}{\sum_{0 \leq k < m} \mathbb{1}_{\{x\}}(X_k)} \to \lambda_x(f) \; \mathbb{P}_x \text{ p.s.} \tag{4.25}$$

Ecrivant $f = f^+ - f^-$, on a (4.25) pour $f \in L^1(\lambda_x)$. Enfin si $f, g \in L^1(\lambda_x)$ et $\lambda_x(g) \neq 0$, on a (4.25) pour f et g et on fait le quotient ∎

4.6. Théorie du potentiel des chaînes de Markov

Soient Q une matrice de transition et $X = (\Omega, \mathcal{F}_n, \mathcal{F}, X_n, \mathbb{P}_x)$ la chaîne canonique associée. Cette section a pour objet de donner des procédés de calcul de quantités telles que $U(x,x)$, $\mathbb{P}_x(S_x < \infty)$, $\mathbb{E}_x(S_x)$ qui sont déterminantes pour l'étude de la nature de la chaîne .

Opérateurs induits

Soient S et T deux temps d'arrêt. On considère $R = T + S \circ \theta_T$. Sur $T = +\infty$, R est défini sans ambiguïté par $R = +\infty$ et on n'est pas gêné par le fait que θ_T ne soit pas défini. Par exemple si T est le temps d'entrée dans A et S le temps d'entrée dans B, R est le premier temps de visite à B après avoir visité A. On a à ce sujet,

Lemme 4.6.1. *Soient S et T deux temps d'arrêt et $R = T + S \circ \theta_T$. Alors R est un temps d'arrêt et, sur $\{R < +\infty\}$, $X_R = X_S \circ \theta_T$.*

`Preuve:`
On a, $\{R = n\} = \cup_{k=0}^{n}\{T = k\} \cap \{S \circ \theta_k = n - k\} \in \mathcal{F}_n$ pour tout n, car $\{S \circ \theta_k = n - k\} = \theta_k^{-1}(\{S = n - k\}) \in \mathcal{F}_n$ vu (4.20). De plus $R < +\infty$ implique $T < +\infty$ et $S \circ \theta_T < +\infty$ et alors

$$X_R(\omega) = X_{T(\omega)+S(\theta_T(\omega))}(\omega) = X_{S(\theta_T(\omega))}(\theta_T(\omega)) = X_S \circ \theta_T(\omega) \ \blacksquare$$

On en conclut par exemple que sur $\{S_A^{n+1} < \infty\}$:

$$X_{S_A^{n+1}} = X_{S_A^n} \circ \theta_{S_A} = X_{S_A} \circ \theta_{S_A^n}$$

Soit T un temps d'arrêt . On définit l'opérateur Q_T de $\mathcal{E}^+$ dans $\mathcal{E}^+$ (ou de $\mathcal{E}_b$ dans $\mathcal{E}_b$) par

$$Q_T f(x) = \mathbb{E}_x(\mathbb{1}_{\{T<+\infty\}} f(X_T)). \tag{4.26}$$

On a alors

Proposition 4.6.2. *Soient X une chaîne de Markov canonique, S et T deux temps d'arrêt. On pose $R = T + S \circ \theta_T$. On a alors $Q_R = Q_T Q_S$.*

`Preuve:`
Soit $f \in \mathcal{E}^+$. Alors, d'après le th.4.4.4 et le lem.4.6.1,

$$\begin{aligned} Q_T Q_S f(x) &= \mathbb{E}_x(\mathbb{1}_{\{T<+\infty\}} Q_S f(X_T)) = \mathbb{E}_x(\mathbb{1}_{\{T<+\infty\}} \mathbb{E}_{X_T}(\mathbb{1}_{\{S<+\infty\}} f(X_S))) \\ &= \mathbb{E}_x(\mathbb{1}_{\{T<+\infty\}} \mathbb{1}_{\{S\circ\theta_T<+\infty\}} f(X_S \circ \theta_T)) \\ &= \mathbb{E}_x(\mathbb{1}_{\{R<+\infty\}} f(X_R)) = Q_R f(x) \ \blacksquare \end{aligned}$$

Calcul de la matrice potentielle

Si $f \in \mathcal{E}^+$, $v = Uf$ s'appelle le potentiel de f. On a, compte tenu de (4.18), $\sum_{k=0}^{\infty} Q^k f(x) = \sum_{k=0}^{\infty} \mathbb{E}_x(f(X_k)) = \mathbb{E}_x\left(\sum_{k=0}^{\infty} f(X_k)\right)$ d'où

$$Uf(x) = \mathbb{E}_x(\sum_{k=0}^{\infty} f(X_k)), \quad f \in \mathcal{E}^+. \tag{4.27}$$

En particulier, si $f = \mathbb{1}_{\{A\}}$, $A \subset E$, on a:

$$U(x, A) = U\mathbb{1}_{\{A\}}(x) = \mathbb{E}_x\left(\sum_{k=0}^{\infty} \mathbb{1}_{\{A\}}(X_k)\right) = \mathbb{E}_x(N_A)$$

De la relation $U = I + QU$ on déduit que $v = f + Qv$.

Proposition 4.6.3. *Pour $f \in \mathcal{E}^+$, $v = Uf$ est la plus petite solution positive de l'équation $v = f + Qv$.*

`Preuve:`
On a $v = f + Qv$. Soit $w \in \mathcal{E}^+$ vérifiant $w = f + Qw$. Evidemment $w \geq f$. Supposons $w \geq \sum_{k=0}^{n} Q^k f$, on a

$$w = f + Qw \geq f + Q \sum_{k=0}^{n} Q^k f = \sum_{k=0}^{n+1} Q^k f.$$

Donc, pour tout n, $w \geq \sum_{k=0}^{n} Q^k f$ et, à la limite, $w \geq v$ ∎

Remarquant que $U\mathbb{1}_{\{y\}}(x) = U(x,y)$, on a

Corollaire 4.6.4. *Pour tout $y \in E$, $v(x) = U(x,y)$ est la plus petite solution positive de l'équation*

$$v(y) = 1 + Qv(y), \quad v(x) = Qv(x), \ \ x \neq y\,.$$

Ceci peut fournir un moyen de calculer la matrice potentielle. Il faut faire attention que par solution positive, on entend solution à valeurs $[0, +\infty]$. En particulier, $v \equiv +\infty$ fait partie des solutions de l'équation considérée et, dans de nombreux cas, ce sera la seule.

Problèmes de passage

On s'intéresse à l'opérateur Q_{T_A} défini par (4.26) et qu'on notera Q_A. On a donc

$$Q_A f(x) = \mathbb{E}_x(\mathbb{1}_{\{T_A < +\infty\}} f(X_{T_A})), \ \ f \in \mathcal{E}^+. \tag{4.28}$$

Remarquons que, si $f = 1$, on a $Q_A 1(x) = \mathbb{P}_x(T_A < +\infty)$ qui est la probabilité pour que la chaîne partant de x atteigne A et que, si $f = \mathbb{1}_{\{y\}}$,

$$Q_A \mathbb{1}_{\{y\}}(x) = \mathbb{P}_x(T_A < +\infty, \, X_{T_A} = y). \tag{4.29}$$

qui est la probabilité pour que la chaîne partant de x pénètre dans A par le point y.

Théorème 4.6.5. *Pour $f \in \mathcal{E}^+$, $u(x) = Q_A f(x)$ est la plus petite solution positive du système:*

$$u(x) = f(x), \ x \in A\,; \quad u(x) = Qu(x), \ x \in A^c\,.$$

`Preuve:`
• Soit $u(x) = Q_A f(x)$. Pour $x \in A$, on a $u(x) = f(x)$ et, pour $x \in A^c$, vu la prop.4.6.2 et la relation $S_A = 1 + T_A \circ \theta$, on a $u(x) = Q_{S_A} f(x) = QQ_A f(x) = Qu(x)$.

• Soit $v \in \mathcal{E}^+$ vérifiant $v(x) = f(x), \ x \in A, \ v(x) = Qv(x), \ x \in A^c$. On pose $T = T_A$ et $Z_n = v(X_{n \wedge T})$. Alors $\mathbb{E}_x(Z_n) = \mathbb{E}_x(Z_0) = v(x)$. Pour montrer ceci, on va en fait prouver beaucoup plus, à savoir que Z_n est une martingale, c'est à dire que $\mathbb{E}_x(Z_{n+1}|\mathcal{F}_n) = Z_n$. On remarque tout d'abord que la quantité $Z_{n+1}\mathbb{1}_{\{T \leq n\}} = Z_n \mathbb{1}_{\{T \leq n\}}$ est $\mathcal{F}_n$ mesurable et donc:

$$\begin{aligned}
Z_{n+1} &= Z_{n+1}\mathbb{1}_{\{T \leq n\}} + v(X_{n+1})\mathbb{1}_{\{T > n\}} \\
\mathbb{E}_x(Z_{n+1}|\mathcal{F}_n) &= Z_{n+1}\mathbb{1}_{\{T \leq n\}} + \mathbb{E}_x(v(X_{n+1})\mathbb{1}_{\{T > n\}}|\mathcal{F}_n) \\
&= Z_{n+1}\mathbb{1}_{\{T \leq n\}} + \mathbb{1}_{\{T > n\}}\mathbb{E}_x(v(X_{n+1})|\mathcal{F}_n) \\
&= Z_n\mathbb{1}_{\{T \leq n\}} + \mathbb{1}_{\{T > n\}} Qv(X_n) = Z_n
\end{aligned}$$

En effet, sur l'ensemble $\{T > n\}$ on a $X_n \in A^c$, et donc $v(X_n) = Qv(X_n)$. La suite Z_n converge vers $v(X_T) = f(X_T)$ sur $\{T < \infty\}$. On a alors:

$$u(x) = \mathbb{E}_x(\mathbb{1}_{\{T<+\infty\}} f(X_T)) = \mathbb{E}_x(\lim_n \mathbb{1}_{\{T<+\infty\}} Z_n) \le \underline{\lim}_n \mathbb{E}_x(Z_n) = v(x) \quad \blacksquare$$

Corollaire 4.6.6. *$u(x) = \mathbb{P}_x(T_A < +\infty)$ est la plus petite solution positive du système:*

$$u(x) = 1,\ x \in A\,; \qquad u(x) = Qu(x),\ x \in A^c\,.$$

et, vu (4.29),

Corollaire 4.6.7. *Pour tout $y \in A$, $u(x) = \mathbb{P}_x(T_A < +\infty,\, X_{T_A} = y)$ est la plus petite solution positive du système:*

$$u(y) = 1\,; \quad u(x) = 0,\ x \in A \setminus \{y\}\,; \quad u(x) = Qu(x),\ x \in A^c\,.$$

On s'intéresse maintenant à $v(x) = \mathbb{E}_x(T_A)$.

Théorème 4.6.8. *$v(x) = \mathbb{E}_x(T_A)$ est la plus petite solution positive de*

$$v(x) = 0,\ x \in A\,; \quad v(x) = 1 + Qv(x),\ x \in A^c\,.$$

`Preuve:`

• Si $x \in A$, $v(x) = 0$. Si $x \in A^c$, on a:

$$\begin{aligned} v(x) &= \mathbb{E}_x(T_A) = \mathbb{E}_x(S_A) = 1 + \mathbb{E}_x(T_A \circ \theta) \\ &= 1 + \mathbb{E}_x(\mathbb{E}_{X_1}(T_A)) = 1 + \mathbb{E}_x(v(X_1)) = 1 + Qv(x) \end{aligned}$$

• Soit $w \in \mathcal{E}^+$ vérifiant $w(x) = 0$, $x \in A$, $w(x) = 1 + Qw(x)$, $x \in A^c$. On pose $T = T_A$ et $Y_n = w(X_{n\wedge T}) + n \wedge T$. Alors $\mathbb{E}_x(Y_n) = \mathbb{E}_x(Y_0) = w(x)$. Pour montrer ceci, on va de nouveau prouver beaucoup plus, à savoir que Y_n est une martingale, c'est à dire que $\mathbb{E}_x(Y_{n+1}|\mathcal{F}_n) = Y_n$. On remarque tout d'abord que la quantité $Y_{n+1}\mathbb{1}_{\{T\le n\}} = Y_n \mathbb{1}_{\{T\le n\}}$ est $\mathcal{F}_n$ mesurable et donc:

$$\begin{aligned} Y_{n+1} &= Y_{n+1}\mathbb{1}_{\{T\le n\}} + (w(X_{n+1}) + (n+1))\mathbb{1}_{\{T>n\}} \\ \mathbb{E}_x(Y_{n+1}|\mathcal{F}_n) &= Y_{n+1}\mathbb{1}_{\{T\le n\}} + \mathbb{E}_x(w(X_{n+1}) + (n+1))\mathbb{1}_{\{T>n\}}|\mathcal{F}_n) \\ &= Y_{n+1}\mathbb{1}_{\{T\le n\}} + \mathbb{1}_{\{T>n\}}\mathbb{E}_x(w(X_{n+1}) + (n+1)|\mathcal{F}_n) \\ &= Y_n\mathbb{1}_{\{T\le n\}} + \mathbb{1}_{\{T>n\}}(Qw(X_n) + (n+1)) = Y_n \end{aligned}$$

En effet, sur l'ensemble $\{T > n\}$ on a $X_n \in A^c$, et donc $w(X_n) = Qw(X_n) + 1$. Or sur $\{T < \infty\}$, la suite Y_n converge vers $w(X_T) + T = T$ et sur $\{T = \infty\}$, la suite Y_n converge vers $+\infty$, donc dans tous les cas cette suite converge vers T. On a par conséquent:

$$\mathbb{E}_x(T) \le \underline{\lim}_n \mathbb{E}_x(Y_n) = \mathbb{E}_x(Y_0) = w(x) \quad \blacksquare$$

Corollaire 4.6.9. *Si l'on a calculé les fonctions $u(x) = \mathbb{P}_x(T_y < \infty)$ et $v(x) = \mathbb{E}_x(T_y)$, par exemple en utilisant le corollaire 4.6.6 ou le théorème 4.6.8, alors on a $\mathbb{P}_y(S_y < \infty) = Qu(y)$ et $\mathbb{E}_y(S_y) = 1 + Qv(y)$.*

`Preuve:`

Il suffit d'utiliser la relation $S_y = 1 + T_y \circ \theta$ et d'appliquer la propriété de Markov. $\blacksquare$

4.7. Chaînes irréductibles récurrentes

On considère une chaîne de Markov canonique X de matrice de transition Q, irréductible et récurrente. On a donc pour tous $x, y \in E$, $\mathbb{P}_x(N_y = \infty) = 1$. Nous allons commencer par l'étude des mesures ρ invariantes (harmoniques) c'est à dire telles que $\rho Q = \rho$ ou excessives (surharmoniques) c'est à dire telles que $\rho Q \leq \rho$.

Définition 4.7.1. *On dit qu'une mesure sur E est triviale si c'est soit la mesure nulle soit la mesure valant $+\infty$ en tout point.*

Ces deux mesures triviales sont considérées comme "parasites" car ce sont toujours des mesures invariantes et leur prise en compte n'offre aucun intérêt.

Lemme 4.7.2. *Soit ρ une mesure excessive non triviale, alors ρ est strictement positive et finie en tout point.*

`Preuve:`
Soient a et b deux points de E. Vu l'irréductibilité, il existe $n \geq 0$ tel que $Q^n(b, a) > 0$. Comme $\rho \geq \rho Q^n$, on a

$$\rho(a) \geq \sum_c \rho(c) Q^n(c, a) \geq \rho(b) Q^n(b, a).$$

On en déduit facilement le résultat ∎

Lemme 4.7.3. *Soit ρ une mesure excessive finie en tout point, alors ρ est invariante.*

`Preuve:`
Puisque $\rho Q(x) \leq \rho(x) < \infty$, on peut considérer la mesure positive $\nu = \rho - \rho Q$. Or on a $\sum_{k=0}^{n} \nu Q^k = \rho - \rho Q^{n+1}$ et donc $\nu U \leq \rho$. Si ν n'est pas la mesure nulle, on a $\nu U(y) = \sum_x U(x, y)\nu(x) = \infty$ car $U(x, y) \equiv \infty$. Ceci est impossible et par conséquent ρ est invariante ∎

Proposition 4.7.4. *On suppose que la chaîne admet une mesure invariante λ non triviale. Alors toute mesure excessive (et donc toute mesure invariante) lui est proportionnelle.*

Soit ρ une mesure excessive. On peut bien sûr supposer ρ non triviale sinon le résultat est évident. Soit alors c un réel positif. La mesure $\nu_c(x) = \inf(\rho(x), c\lambda(x))$ est excessive et finie. Par conséquent elle est invariante et il en est de même de la mesure positive $\mu_c = c\lambda - \nu_c$. Il y a donc deux cas: ou bien μ_c est la mesure nulle ou bien elle est partout strictement positive. Dans le premier cas la fonction $f(x) = \rho(x)/\lambda(x)$ (qui est bien définie) est supérieure ou égale à c, dans le second cas elle est strictement inférieure à c. Cette alternative ayant lieu pour tout $c \geq 0$ on en déduit que f est constante ∎

Théorème 4.7.5. *Soit X une chaîne irréductible récurrente. Il existe une mesure invariante λ non triviale, unique à une constante multiplicative près. De plus, pour tout $y \in E$, $0 < \lambda(y) < +\infty$. Pour chaque $x \in E$, la mesure invariante telle que $\lambda(x) = 1$ est donnée par la mesure λ_x définie en (4.5.17).*

`Preuve:`
Il suffit de montrer que la mesure λ_x définie par (4.5.17) est invariante. Soit $f \in \mathcal{E}^+$:

$$\begin{aligned}
(\lambda_x Q)f &= \lambda_x(Qf) = \mathbb{E}_x(\sum_{k=0}^{S_x-1} Qf(X_k)) \\
&= \sum_{k\geq 0} \mathbb{E}_x(\mathbb{1}_{\{k<S_x\}} Qf(X_k)) = \sum_{k\geq 0} \mathbb{E}_x(\mathbb{1}_{\{k<S_x\}} \mathbb{E}_x(f(X_{k+1}|\mathcal{F}_k))) \\
&= \sum_{k\geq 0} \mathbb{E}_x(\mathbb{1}_{\{k<S_x\}} f(X_{k+1})) = \mathbb{E}_x(\sum_{k=1}^{S_x} f(X_k)) \\
&= \lambda_x(f) - \mathbb{E}_x(f(X_0)) + \mathbb{E}_x(f(X_{S_x})) = \lambda_x(f)
\end{aligned}$$

Puisque $\lambda_x(x) = 1$, λ est donc non triviale et par conséquent strictement positive et finie en tout point ∎

Il y a une nouvelle dichotomie selon que $\lambda(E) = +\infty$ ou $\lambda(E) < +\infty$.

Théorème 4.7.6. *Soit X une chaîne irréductible récurrente de mesure invariante λ. Il y a deux cas possibles:*

- $\lambda(E) = +\infty$. *Alors, pour tout $x \in E$, $\mathbb{E}_x(S_x) = +\infty$. La chaîne est dite récurrente nulle.*
- $\lambda(E) < +\infty$. *Alors, pour tout $x \in E$, $\mathbb{E}_x(S_x) < +\infty$. La chaîne est dite récurrente positive. Dans ce cas l'unique probabilité invariante est donnée par*

$$\pi(x) = 1/\mathbb{E}_x(S_x) \tag{4.30}$$

`Preuve:`
Vu (4.5.17), on a, dans le premier cas, $\mathbb{E}_x(S_x) = +\infty$ et dans le second cas, $\mathbb{E}_x(S_x) < +\infty$. Enfin si $\lambda(E) < +\infty$, on a

$$\pi(x) = \frac{\lambda_x(x)}{\lambda_x(E)} = \frac{\lambda_x(x)}{\mathbb{E}_x(S_x)} = \frac{1}{\mathbb{E}_x(S_x)} \quad ∎$$

Pour calculer la mesure invariante de la chaîne, on n'utilise en général pas (4.5.17). On résout directement l'équation $\lambda = \lambda Q$. Cependant cette équation peut être difficile à résoudre d'où l'intérêt de la notion de mesure réversible introduite dans 4.3.7.

Le théorème ergodique dans le cas irréductible

Théorème 4.7.7. *Soit X une chaîne irréductible récurrente de mesure invariante λ. Soient $f, g \in L^1(\lambda)$ avec $\lambda(g) \neq 0$, alors pour toute loi initiale μ on a:*

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \to \frac{\lambda(f)}{\lambda(g)} \quad \mathbb{P}_\mu \ p.s.$$

`Preuve:`
Soit x un point fixé de E. D'après le théorème 4.5.18, les quotients ci dessus convergent $\mathbb{P}_x$ p.s. vers $\frac{\lambda_x(f)}{\lambda_x(g)}$ si f et g sont dans $L^1(\lambda_x)$ avec $\lambda_x(g) \neq 0$. Mais puisque λ est proportionnelle à λ_x avec une constante de proportionnalité strictement comprise entre 0 et $+\infty$, on voit que la limite ne dépend pas de x et que les conditions d'intégrabilité sont satisfaites pour λ. Il s'en suit que l'ensemble Ω_0 sur lequel la convergence des quotients vers $\frac{\lambda(f)}{\lambda(g)}$ a lieu, est tel que $\mathbb{P}_x(\Omega_0) = 1$ pour tout $x \in E$ et donc $\mathbb{P}_\mu(\Omega_0) = \int \mathbb{P}_x(\Omega_0)\, d\mu(x) = 1$ ∎

Notation: Pour simplifier l'écriture, on notera "$Z_n \to Z$ p.s." pour " $Z_n \to Z$ $\mathbb{P}_\mu$ p.s. pour toute loi initiale μ".

Corollaire 4.7.8. *Si on choisit $f = \mathbb{1}_{\{y\}}$ et $g = \mathbb{1}_{\{x\}}$, on a,*

$$\frac{\textit{nombre de visites de } X \textit{ à } y \textit{ avant } n}{\textit{nombre de visites de } X \textit{ à } x \textit{ avant } n} \to \frac{\lambda(y)}{\lambda(x)} \quad \textit{p.s.}$$

Le théorème suivant explique le choix des termes "récurrent positif" et "récurrent nul".

Théorème 4.7.9. *Soit X une chaîne irréductible récurrente.*

1. *Dans le cas récurrent positif, si on note π la probabilité invariante, pour toute $f \in L^1(\pi)$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \pi(f) \quad \textit{p.s.}$$

2. *Dans le cas récurrent nul, si on note λ la mesure invariante, pour toute $f \in L^1(\lambda)$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to 0 \quad \textit{p.s.}$$

`Preuve:`
• 1. Il suffit de prendre $g \equiv 1$ dans le th.4.7.7.

• 2. On prend $g = \mathbb{1}_{\{F\}}$ dans le th.4.7.7 avec F fini $\subset E$, on a (en supposant f positive):

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \leq \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} \mathbb{1}_{\{F\}}(X_k)} \to \frac{\lambda(f)}{\lambda(F)}, \quad \text{p.s.}$$

et donc

$$\overline{\lim}\ \frac{1}{n}\sum_{k=0}^{n} f(X_k) \leq \frac{\lambda(f)}{\lambda(F)} \quad \text{p.s.}$$

pour tout F fini d'où le résultat puisque $\lambda(E) = +\infty$ ∎

4.8. Stabilisation des chaînes de Markov

Définition 4.8.1. *On dit qu'une chaîne de Markov se stabilise, s'il existe une probabilité π sur E telle que pour tout point $y \in E$ et toute loi initiale μ on ait:*

$$\lim_{n\to\infty} \mathbb{P}_\mu(X_n = y) = \lim_{n\to\infty} \mu Q^n(y) = \pi(y)$$

En fait, puisque $Q^n(x,y) \leq 1$ et $\mu Q^n(y) = \sum_{x\in E} Q^n(x,y)\mu(x)$ une simple application du théorème de Lebesgue montre que la stabilisation est équivalente à la convergence de $Q^n(x,y)$ vers $\pi(y)$ pour tout $x \in E$. On remarque aussi que π est nécessairement invariante puisque $\mu Q^{n+1}(y) = \sum_{x\in E} \mu Q^n(x)Q(x,y)$ et de nouveau le théorème de Lebesgue permet d'affirmer que $\mu Q^{n+1}(y)$ converge vers $\pi Q(y)$ mais cette quantité converge aussi vers $\pi(y)$. Cette propriété de stabilisation est très importante car elle signifie que pour n "grand" on peut approximer la loi de X_n par la probabilité π, et ceci quelles que soient les conditions initiales.

Dans le cas récurrent irréductible, les fractions figurant dans le thm .4.7.9 pour $f = \mathbb{1}_{\{x\}}$ sont bornées par 1, on peut donc prendre l'espérance $\mathbb{E}_y$ et appliquer le théorème de Lebesgue. Vu que $\mathbb{E}_y(\mathbb{1}_{\{x\}}(X_k)) = Q^k(y,x)$, on obtient,
(i) si X est récurrente positive de probabilité invariante π,

$$\frac{1}{n}\sum_{k=0}^{n-1} Q^k(y,x) \to \pi(x),$$

(ii) si X est récurrente nulle,

$$\frac{1}{n}\sum_{k=0}^{n-1} Q^k(y,x) \to 0.$$

Dans le cas d'une chaîne transitoire, on a bien sûr $\lim_n Q^n(y,x) = 0$ puisque la série est convergente. Ceci montre que le seul cas où l'on peut espérer une limite non triviale pour la suite $Q^n(x,y)$ est le cas récurrent positif et c'est donc le seul qui va nous intéresser ici. Malheureusement même dans ce cas, une telle limite n'existe pas toujours comme le montre la chaîne (déterministe!) à valeurs $E = \{1,2\}$ de matrice de transition $Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. En effet $Q^{2n} = I$ et $Q^{2n+1} = Q$. On voit ici apparaître un phénomène de périodicité. En fait ce problème est déjà celui qui est rencontré dans l'étude des suites récurrentes de type $u_{n+1} = F(u_n)$. Il se peut très bien que l'équation $u = F(u)$ admette une solution unique sans que la suite u_n converge. Dans notre cas, on choisit une mesure initiale μ et on définit la suite $\mu_{n+1} = \mu_n Q = F(\mu_n)$. L'existence et l'unicité de la solution de $\mu Q = \mu$ dans $\mathcal{M}_1(E)$ ne garantit pas la convergence de la suite μ_n vers cette solution (voir par contre le cas des chaînes de Doeblin thm. 4.8.11).

Soient Q une matrice de transition et X la chaîne canonique de matrice de transition Q. Pour $a \in E$, on pose

$$I(a) = \{n > 0,\ Q^n(a,a) > 0\}. \tag{4.31}$$

Définition 4.8.2. *Lorsque $I(a)$ est non vide, on appelle période de a et on note d_a le pgcd de l'ensemble $I(a)$. Si tous les états ont une période et que celle-ci vaut 1, la chaîne est dite apériodique.*

Proposition 4.8.3. *Si X est une chaîne irréductible, tous les états ont une période et c'est la même.*

`Preuve:`
Soient $x \neq y$. Puisque $y \to x$ et $x \to y$ on a $I(y) \neq \emptyset$ et pour $n \in I(y)$, on a $Q^n(y,y) > 0$. Comme $x \leftrightarrow y$, il existe n_1 et n_2 tels que $Q^{n_1}(x,y) > 0$ et $Q^{n_2}(y,x) > 0$. Alors $Q^{n_1+n_2}(x,x) \geq Q^{n_1}(x,y)Q^{n_2}(y,x) > 0$ et $Q^{n_1+n+n_2}(x,x) \geq Q^{n_1}(x,y)Q^n(y,y)Q^{n_2}(y,x) > 0$ donc d_x divise n_1+n_2 et n_1+n+n_2 et donc n d'où $d_x \leq d_y$. Par symétrie $d_y \leq d_x$ et $d_x = d_y$ ∎

Remarque: Il résulte de la prop.4.8.3 que, si X est irréductible, elle est apériodique dès qu'un état est de période 1, en particulier dès qu'il existe $a \in E$ tel que $Q(a,a) > 0$.

Lemme 4.8.4. *Soit $A \subset \mathbb{N}^*$ un ensemble stable par addition et de p.g.c.d. égal à 1. Alors il existe N tel que A contienne tous les entiers $\geq N$.*

Soit $A' = A \cup \{0\}$. L'ensemble $(A' - A')$ est un sous groupe de $\mathbb{Z}$ qui est donc de la forme $d\mathbb{Z}$ où d est le plus petit élément non nul de $(A'-A')$ et de plus $(A'-A')$ contient A. Tout élément de A étant divisible par d on en conclut que $d = 1$ et que donc il existe a et b dans A' avec $a = b+1$. On a nécessairement $a \in A$ et si $b = 0$ la preuve est terminée, on peut donc supposer $b \in A$. Alors l'entier $N = b^2$ convient puisque si $n \geq N$ on peut écrire $n = b^2 + bq + r$ avec $0 \leq r < b$ et donc $n = b(b+q-r) + r(b+1) \in A$ ∎

Proposition 4.8.5. *Supposons X irréductible et apériodique. Alors pour tous $x, y \in E$, il existe N tel que, pour tout $n \geq N$, $Q^n(x,y) > 0$.*

`Preuve:`
Soit $a \in E$. Il est immédiat que, si $p, q \in I(a)$, $p+q \in I(a)$. par conséquent, d'après 4.8.4, il existe n_1 tel que $I(a)$ contienne tous les entiers $\geq n_1$. Considérons maintenant $x, y \in E$. Il existe n_2 et n_3 tels que $Q^{n_2}(x,a) > 0$, $Q^{n_3}(a,y) > 0$ et pour $n \geq n_1$, on a

$$Q^{n_2+n+n_3}(x,y) \geq Q^{n_2}(x,a)Q^n(a,a)Q^{n_3}(a,y) > 0$$

On choisit donc $N = n_1 + n_2 + n_3$ ∎

Remarque 4.8.6. *Il faut noter que, dans la prop.4.8.5, l'entier N dépend des états x et y. Cependant elle montre que, si E est fini, il existe N tel que, pour tout $n \geq N$, la matrice Q^n ait tous ses termes strictement positifs.*

Définition 4.8.7. *Soit $X = (\Omega, X_n, \mathbb{P}_\alpha)$ une chaîne canonique et soit $X' = (\Omega', X'_n, \mathbb{P}'_\beta)$ une copie indépendante de X. On définit la chaîne "couplée" Z par:*

$$Z = (\Omega \times \Omega', (X_n, X'_n), \mathbb{P}_{\alpha\otimes\beta}) \ \text{ où } \mathbb{P}_{\alpha\otimes\beta} = \mathbb{P}_\alpha \otimes \mathbb{P}'_\beta$$

On vérifie facilement que $Z_n = (X_n, X'_n)$ est une chaîne de Markov à valeurs dans $E \times E$, de matrice de transition $\bar{Q}$ donnée par

$$\bar{Q}((x,x'),(y,y')) = Q(x,y)Q(x',y')$$

Corollaire 4.8.8. *Soit X une chaîne irréductible, récurrente positive et apériodique. Alors la chaîne couplée est récurrente positive irréductible.*

`Preuve:`
On vérifie facilement que $\bar{Q}^n((x,x'),(y,y')) = Q^n(x,y)Q^n(x',y')$. On a donc (prop.4.8.5) $\bar{Q}^n((x,x'),(y,y')) > 0$ pour tout n assez grand, ce qui montre que la chaîne Z est irréductible. De plus si π est la probabilité invariante de X alors $\pi \otimes \pi$ est une probabilité invariante pour Z. Il ne reste plus qu'à utiliser la prop. 4.5.16 ∎

Le résultat fondamental est:

Théorème 4.8.9. *Soit X une chaîne irréductible récurrente positive apériodique de probabilité invariante π. Alors cette chaîne se stabilise. Plus précisément:*

1. *Pour toutes probabilités α, β sur E:*

$$\sup_{y \in E} |\alpha Q^n(y) - \beta Q^n(y)| \to_n 0$$

2. *Il s'en suit que pour tous $x, y \in E$,*

$$Q^n(x,y) \to_n \pi(y)$$

`Preuve:`
On choisit un point fixé $a \in E$ et on pose $R = \inf\{n > 0 \ ; \ Z_n = (a,a)\}$. Puisque la chaîne Z_n est récurrente irréductible on sait que $\mathbb{P}_{\alpha\otimes\beta}(R < \infty) = 1$. Pour deux fonctions f et g dans $\mathcal{E}_b$, on a:

$$\mathbb{E}_{(\alpha\otimes\beta)}\big(f(X_n)g(X'_n)\big) = \mathbb{E}_\alpha\big(f(X_n)\big)\mathbb{E}_\beta\big(g(X'_n)\big) = \big(\alpha Q^n(f)\big)\big(\beta Q^n(g)\big)$$

Par conséquent:

$$\begin{aligned}
\alpha Q^n(f) &= \mathbb{E}_{(\alpha\otimes\beta)} f(X_n) \\
&= \mathbb{E}_{(\alpha\otimes\beta)}(f(X_n)\mathbb{1}_{\{R>n\}}) + \sum_{k=1}^n \mathbb{E}_{(\alpha\otimes\beta)}(f(X_n)\mathbb{1}_{\{R=k\}}) \\
&= \mathbb{E}_{(\alpha\otimes\beta)}(f(X_n)\mathbb{1}_{\{R>n\}}) + \sum_{k=1}^n \mathbb{E}_{(\alpha\otimes\beta)}(\mathbb{1}_{\{R=k\}})\mathbb{E}_{(a,a)} f(X_{n-k})) \\
&= \mathbb{E}_{(\alpha\otimes\beta)}(f(X_n)\mathbb{1}_{\{R>n\}}) + \sum_{k=1}^n Q^{n-k} f(a)\mathbb{P}_{(\alpha\otimes\beta)}(R = k)
\end{aligned}$$

La troisième ligne est obtenue en écrivant $X_n = X_{n-k} \circ \theta_k$ et en appliquant la propriété de Markov. On en déduit que $|\alpha Q^n(f) - \beta Q^n(f)| \leq 2||f||\mathbb{P}_{(\alpha\otimes\beta)}(R > n)$ et donc en choisissant $f = \mathbb{1}_{\{y\}}$, l'inégalité fondamentale:

$$|\alpha Q^n(y) - \beta Q^n(y)| \leq 2\mathbb{P}_{(\alpha\otimes\beta)}(R > n)$$

ce qui prouve 1. Pour obtenir la seconde assertion il suffit de choisir $\beta = \pi$ ∎

Chaînes de Doeblin

Vu leur importance, nous donnons une seconde démonstration des th.4.7.5 et 4.8.9 dans le cas particulier des chaînes de Doeblin (qui concerne essentiellement les espaces d'états finis). Elle a de plus l'avantage de préciser la vitesse de convergence.

Définition 4.8.10. *On dit qu'une probabilité de transition* **Q** *vérifie la condition de Doeblin s'il existe une probabilité ν sur E et un réel $p > 0$ avec*

$$Q(x, y) \geq p\nu(y), \quad \textit{pour tout} \quad y \in E$$

La constante p est nécessairement ≤ 1 et on supposera dans la suite que l'on est pas dans le cas trivial $Q = \nu$ autrement dit que l'on ne peut pas choisir $p = 1$. Il existe alors une probabilité de transition S telle qu'en posant $q = 1 - p$ on ait:

$$Q(x, y) = p\nu(y) + qS(x, y)$$

On peut identifier l'espace des mesures signées et bornées sur E, soit $\mathcal{M}_b(E)$, à l'espace $\ell^1(E)$ en utilisant la norme: $||\mu|| = \sum_{x\in E} |\mu(x)|$ Cet espace de mesures est alors un espace de Banach réel et le sous ensemble $\mathcal{M}_1(E)$ est fermé, donc lui-même complet.

Théorème 4.8.11. *Soit Q une probabilité de transition vérifiant la condition de Doeblin. Il existe alors une unique probabilité invariante λ et pour toute probabilité μ sur E:*

$$||\mu Q^n - \lambda|| \leq 2q^n$$

De plus, la probabilité invariante est donnée par la formule $\lambda = p \sum_{n=0}^{\infty} q^n \nu S^n$

`Preuve:`
Soit $\rho \in \mathcal{M}_b(E)$ de masse totale nulle:

$$\begin{aligned} \rho Q(y) &= \sum_x \rho(x)(p\nu(y) + qS(x, y)) = q \sum_x \rho(x)S(x, y) \\ ||\rho Q|| &= \sum_y |\rho Q(y)| \leq q \sum_x \left(|\rho(x)|(\sum_y S(x, y)) \right) = q||\rho|| \end{aligned}$$

Par conséquent pour μ_1 et μ_2 dans $\mathcal{M}_1(E)$ on a $||\mu_1 Q - \mu_2 Q|| \leq q||\mu_1 - \mu_2||$ L'application $\mu \to F(\mu) := \mu Q$ est donc globalement contractante sur l'espace complet $\mathcal{M}_1(E)$, d'où l'unicité du point fixe et la convergence de la suite $\mu_n = \mu Q^n$ à la vitesse q^n vers ce point fixe pour toute probabilité μ. Il reste à montrer que ce point fixe, soit λ, est bien donné par la formule ci dessus, or une probabilité λ est invariante si et seulement si $\lambda = p\nu + q\lambda S$ ce que l'on vérifie immédiatement sur la formule proposée ∎

Corollaire 4.8.12. *Soit* **Q** *une probabilité de transition et* k *un entier* ≥ 1 *tels que* Q^k *vérifie la condition de Doeblin. Il existe alors une unique probabilité invariante* λ*. De plus il existe des constantes* $C < \infty$ *et* $0 < \rho < 1$ *tels que:*

$$||\mu Q^n - \lambda|| \leq C\rho^n, \quad \textit{pour toute probabilité} \quad \mu$$

`Preuve:`
On pose $R = Q^k$. Il existe donc une probabilité R invariante λ avec $||\mu R^n - \lambda|| \leq 2q^n$ pour toute probabilité μ. Si on choisit $\mu = \lambda Q$ alors

$$\mu R^n = \lambda Q^{nk+1} = (\lambda R^n)Q = \lambda Q$$

et l'on obtient donc que λ est Q invariante. Il suffit ensuite d'écrire que si $n = ks + r$ avec $0 \leq r \leq k-1$ on a:

$$||\mu Q^n - \lambda|| = ||\mu Q^{sk+r} - \lambda Q^r|| \leq ||\mu Q^{sk} - \lambda|| \leq 2q^s \leq C\rho^n$$

avec $\rho = q^{1/k}$ et $C = 2/q$ ∎

Remarque 4.8.13. *La convergence ci dessus peut aussi s'interpréter comme suit:*

- *En prenant* $\mu = \epsilon_x$*, on obtient* $|Q^n(x,y) - \lambda(y)| \leq C\rho^n$*, pout tout couple* $(x,y) \in E$
- *Pour toute fonction* f *bornée sur* E *on a* $|Q^n f(x) - \lambda(f)| \leq ||f||_\infty C\rho^n$

Les hypothèses de ce corollaire sont en particulier vérifiées sur un espace d'états fini lorsqu'il existe une puissance de la matrice de transition possédant une colonne strictement positive et donc pour une chaîne irréductible apériodique (voir 4.8.6).

Un exemple typique de chaîne de Doeblin dans le cas d'un espace d'états infini est le suivant:

Pour $n \geq 1$ soit (A_n, B_n, C_n) des variables aléatoires toutes indépendantes entre elles et telles que:

- Les variables aléatoires A_n prennent les valeurs 1 et 0 avec les probabilités p et $1-p$, avec $0 < p < 1$.
- Les variables aléatoires B_n sont à valeurs dans un espace F et de loi μ.
- Les variables aléatoires C_n sont à valeurs dans un espace E et de loi ν.

On se donne de plus une fonction f mesurable de $E \times F$ dans E et on considère la suite X_n de variables aléatoires à valeurs dans E vérifiant:

$$\begin{cases} X_{n+1} = C_{n+1} & \text{si } A_{n+1} = 1 \\ X_{n+1} = f(X_n, B_{n+1}) & \text{si } A_{n+1} = 0 \end{cases}$$

On note S la fonction de transition de la chaîne Z définie par $Z_{n+1} = f(Z_n, B_{n+1})$. Alors X_n est une chaîne de Markov de transition $Q = p\nu + qS$.

4.9. Annexe

4.9.1. Récurrence et fonctions excessives

Définition 4.9.1. *Une fonction $f \in \mathcal{E}^+$ est dite surharmonique ou excessive, si, pour tout x, $f(x) \geq Qf(x)$. Une fonction $f \in \mathcal{E}^+$ est dite harmonique ou invariante, si, pour tout x, $f(x) = Qf(x)$.*

Une classe importante de fonctions excessives est fournie par les potentiels.

Proposition 4.9.2. *Soit X une chaîne irréductible récurrente. Alors toute fonction excessive est constante.*

`Preuve:`
Soit Q la matrice de transition et λ une mesure invariante strictement positive et finie en tous points. On pose

$$\widehat{Q}(x,y) = \frac{\lambda(y)}{\lambda(x)} Q(y,x),$$

$\widehat{Q}$ est une matrice de transition car

$$\sum_y \widehat{Q}(x,y) = \frac{1}{\lambda(x)} \sum_y \lambda(y) Q(y,x) = \frac{1}{\lambda(x)} \lambda(x) = 1.$$

On a alors

$$\widehat{Q}^2(x,y) = \sum_z \widehat{Q}(x,z)\widehat{Q}(z,y) = \sum_z \frac{\lambda(z)}{\lambda(x)} Q(z,x) \frac{\lambda(y)}{\lambda(z)} Q(y,z) = \frac{\lambda(y)}{\lambda(x)} Q^2(y,x)$$

et de même

$$\widehat{Q}^n(x,y) = \frac{\lambda(y)}{\lambda(x)} Q^n(y,x), \qquad \widehat{U}(x,y) = \frac{\lambda(y)}{\lambda(x)} U(y,x).$$

Ceci implique que $\widehat{U}(x,y) = +\infty$ et donc la chaîne de matrice de transition $\widehat{Q}$ est irréductible récurrente. Soit f une fonction excessive et $\rho(x) = f(x)\lambda(x)$. La mesure ρ est excessive pour $\widehat{Q}$ car

$$\rho\widehat{Q}(y) = \sum_x \widehat{Q}(x,y)\lambda(x)f(x) = \lambda(y) \sum_x f(x)Q(y,x) \leq \lambda(y)f(y) = \rho(y)$$

Le même calcul montre que λ est invariante pour $\widehat{Q}$ et donc (prop.4.7.4) $\rho(x)$ est proportionnelle à $\lambda(x)$ i.e. f est constante ∎

Proposition 4.9.3. *Soit X une chaîne de Markov irréductible. X est récurrente ssi toute fonction excessive est constante.*

`Preuve:`
Supposons que toute fonction excessive soit constante. Soient $y \in E$ et $f(x) = \mathbb{P}_x(T_y < +\infty)$. Vu que $\{T_y \circ \theta_1 < +\infty\} \subset \{T_y < +\infty\}$, on a

$$Qf(x) = \mathbb{E}_x(\mathbb{P}_{X_1}(T_y < +\infty)) = \mathbb{P}_x(T_y \circ \theta_1 < +\infty) \leq \mathbb{P}_x(T_y < +\infty) = f(x)$$

i.e. f est excessive et donc constante. Comme $f(y) = 1$, $f \equiv 1$ et X est récurrente car en utilisant la relation $S_x = 1 + T_x \circ \theta$ on a:

$$\mathbb{P}_x(S_x < \infty) = \mathbb{P}_x(T_x \circ \theta < \infty) = \mathbb{E}_x(\mathbb{P}_{X_1}(T_x < \infty)) = 1 \quad \blacksquare$$

4.9.2. Etude d'exemples

Problèmes de passage

On considère la chaîne de Markov à valeurs $\mathbb{Z}$ de matrice de transition

$$Q(x, x+1) = p, \;\; Q(x, x-1) = q$$

où $p, q > 0$ et $p + q = 1$. On pose pour $a, b \in \mathbb{Z}$, $a < b$, et $C \subset \mathbb{Z}$,

$$T_C = \inf(n \geq 0, \; X_n \in C), \;\; T = T_a \wedge T_b.$$

On veut calculer, pour $a < x < b$, $\mathbb{P}_x(T < +\infty)$, $\mathbb{P}_x(T_a < T_b)$ et $\mathbb{P}_x(T_b < T_a)$. D'après le cor.4.6.6, $u(x) = \mathbb{P}_x(T < +\infty) = \mathbb{P}_x(T_{]a,b[^c} < +\infty)$ est solution de

$$u(a) = 1\,; \;\; u(b) = 1\,; \;\; u(x) = Qu(x), \; a < x < b.$$

On a donc pour $a < x < b$, $u(x) = 1 + pu(x+1) + qu(x-1)$ d'où l'on tire

$$u(x+1) - u(x) = \frac{q}{p}(u(x) - u(x-1)) = \ldots = (\frac{q}{p})^{x-a}(u(a+1) - u(a)).$$

Ecrivant $u(x) - u(a) = \sum_{y=a}^{y=x-1}(u(y+1) - u(y))$, on obtient que, pour $x > a$,

$$u(x) = u(a) + \delta\phi_a(x), \; \delta = u(a+1) - u(a), \; \phi_a(x) = 1 + \frac{q}{p} + \ldots (\frac{q}{p})^{x-a-1}. \quad (4.32)$$

Puisque $u(a) = 1$ et $u(b) = 1 = 1 + \delta\phi_a(b)$, on a $\delta = 0$ et $u(x) = \mathbb{P}_x(T < +\infty) = 1$.

D'après le cor.4.6.7, $v(x) = \mathbb{P}_x(T_a < T_b) = \mathbb{P}_x(X_{T_{]a,b[^c}} = a)$ est solution de

$$v(a) = 1\,; \;\; v(b) = 0\,; \;\; v(x) = Qv(x), \; a < x < b.$$

Vu (4.32), on a $v(x) = v(a) + \delta\phi_a(x)$ et $v(b) = 0 = 1 + \delta\phi_a(b)$ d'où $\delta = -\frac{1}{\phi_a(b)}$ et

$$v(x) = \frac{\phi_a(b) - \phi_a(x)}{\phi_a(b)}.$$

Comme $\mathbb{P}_x(T_a < T_b) + \mathbb{P}_x(T_b < T_a) = 1$, on obtient finalement, pour $a < x < b$ et $p \neq q$,

$$\mathbb{P}_x(T_a < T_b) = \frac{(\frac{q}{p})^{b-a} - (\frac{q}{p})^{x-a}}{(\frac{q}{p})^{b-a} - 1}\,, \quad \mathbb{P}_x(T_b < T_a) = \frac{(\frac{q}{p})^{x-a} - 1}{(\frac{q}{p})^{b-a} - 1}.$$

Si $p = q$, on obtient

$$\mathbb{P}_x(T_a < T_b) = \frac{b-x}{b-a}\,, \quad \mathbb{P}_x(T_b < T_a) = \frac{x-a}{b-a}$$

c'est à dire les formules cherchées.

Processus de naissance et de mort

Il s'agit de la chaîne de Markov à valeurs $\mathbb{N}$ de matrice de transition

$$Q(0,1) = p_0, \quad Q(0,0) = r_0,$$
$$Q(x,x+1) = p_x, \quad Q(x,x) = r_x, \quad Q(x,x-1) = q_x, \quad x \geq 1\,,$$

où $p_0 > 0$, $r_0 \geq 0$, $p_0 + r_0 = 1$ et, pour $x \geq 1$, $p_x, q_x > 0$, $r_x \geq 0$, $p_x + r_x + q_x = 1$. On pose, pour $a \in \mathbb{N}$,

$$\gamma_a(a) = 1, \quad \gamma_a(x) = \frac{q_{a+1} \cdots q_x}{p_{a+1} \cdots p_x}, \quad x > a.$$

La chaîne est évidemment irréductible. Pour étudier sa nature, il suffit de calculer $U(0,0)$. Mais (cor.4.6.4) $u(x) = U(x,0)$ est la plus petite solution positive de

$$u(0) = 1 + Qu(0), \quad u(x) = Qu(x), \quad x \geq 1.$$

Cherchons d'abord une solution telle que $u(0) < \infty$. Dans ce cas, d'après le principe du maximum (4.5.2), $u(x)$ est finie en tous points. On a donc $u(0) = 1 + p_0 u(1) + r_0 u(0)$ d'où l'on tire $u(1) - u(0) = -\frac{1}{p_0}$. On a alors, pour $x \geq 1$, $u(x) = p_x u(x+1) + r_x u(x) + q_x u(x-1)$ i.e.

$$u(x+1) - u(x) = \frac{q_x}{p_x}(u(x) - u(x-1)) = \ldots = -\frac{1}{p_0}\frac{q_1 \ldots q_x}{p_1 \ldots p_x} = -\frac{1}{p_0}\gamma_0(x).$$

On en déduit, pour $x \geq 1$,

$$u(x) - u(0) = -\frac{1}{p_0}\sum_{y=0}^{x-1}(u(y+1) - u(y)) = -\frac{1}{p_0}\sum_{y=0}^{x}\gamma_0(y).$$

• Si $\sum_y \gamma_0(y) < +\infty$, $u(x) \to u(+\infty)$ lorsque $x \to +\infty$ et la plus petite solution positive s'obtient en choisissant $u(+\infty) = 0$ ce qui donne $u(0) = \frac{1}{p_0}\sum_y \gamma_0(y)$.

• Si $\sum_y \gamma_0(y) = +\infty$, $u(x) \to -\infty$ lorsque $x \to +\infty$ et il n'y a pas de solution avec $u(0) < +\infty$. On a donc dans ce cas $u(0) = +\infty$ ce qui implique $u(x) = +\infty$ pour tout x, d'après l'équation (4.5.2) et l'irréductibilité.

On a donc

$$U(x,0) \equiv +\infty \text{ si } \sum_y \gamma_0(y) = +\infty, \; U(x,0) = \frac{1}{p_0}\sum_{y=x}^{+\infty} \gamma_0(y) \text{ si } \sum_y \gamma_0(y) < +\infty.$$

On en déduit que

$$\text{si } \sum_y \gamma_0(y) = +\infty, \; X \text{ est récurrente, si } \sum_y \gamma_0(y) < +\infty, \; X \text{ est transiente.}$$

Remarquons que, pour tout $a > 0$,

$$\sum_y \gamma_0(y) = 1 + \frac{q_1}{p_1} + \ldots + \frac{q_1 \ldots q_{a-1}}{p_1 \ldots p_{a-1}} + \frac{q_1 \ldots q_a}{p_1 \ldots p_a} \sum_{y=a}^{+\infty} \gamma_a(y),$$

ce qui montre que la nature de la chaîne ne dépend que de $(p_x, q_x, \; x \geq a)$. Enfin vu que, pour $x > 0$, $U(x,0) = \mathbb{P}_x(T_0 < +\infty)U(0,0)$ (th.4.5.2), on a,

$$\mathbb{P}_x(T_0 < +\infty) = \frac{\sum_{y=x}^{+\infty} \gamma_0(y)}{\sum_{y=0}^{+\infty} \gamma_0(y)}, \; x > 0.$$

Passons à l'étude des mesures invariantes. On commence par chercher une mesure réversible (def.4.3.7) i.e. vérifiant pour tous $x, y \in \mathbb{N}$, $\mu(x)Q(x,y) = \mu(y)Q(y,x)$. Dans ce cas, ceci se réduit à $\mu(x)p_x = \mu(x+1)q_{x+1}$ d'où

$$\mu(x) = \frac{p_0 \ldots p_{x-1}}{q_1 \ldots q_x}\mu(0).$$

Il existe donc toujours une mesure invariante et il existe une probabilité stationnaire ssi

$$\sum_{x=1}^{+\infty} \frac{p_0 \ldots p_{x-1}}{q_1 \ldots q_x} < +\infty.$$

Un cas particulier

Supposons que, pour tout $x \geq 1$, $p_x = p$, $q_x = q$. Alors $\gamma_0(x) = (\frac{q}{p})^x$. La chaîne est transiente si $p > q$, récurrente si $p \leq q$. Elle admet une loi stationnaire, qui est réversible, si $\sum (\frac{p}{q})^x < +\infty$ i.e. si $p < q$.

4.9.3. Marches aléatoires sur $\mathbb{Z}^d$

Soit Y_n une suite de v.a. indépendantes à valeurs $\mathbb{Z}^d$ de même loi ν définies sur $(W, \mathcal{G}, \mathbb{Q})$. On pose

$$S_0 = 0, \;\; S_n = Y_1 + \ldots + Y_n, \;\; n \geq 1.$$

On sait (4.3.5) que $(S_n, \mathbb{Q})$ est une chaîne de Markov de matrice de transition

$$Q(x,y) = \nu(y-x).$$

Le processus $(S_n, \mathbb{Q})$ a donc même loi que la chaîne canonique $(X_n, \mathbb{P}_0)$ de matrice de transition Q. Si on pose $\Pi = \sum_{k=0}^{\infty} \nu^{*k}$ on voit que $U(x,y) = \Pi(y-x)$ et par conséquent $U(x,x) = U(0,0) = \Pi(0)$. Il s'en suit que tous les points de $\mathbb{Z}^d$ sont de même nature et qu'il suffit donc de déterminer si 0 est récurrent ou non pour avoir la réponse pour tous les autres points. Il faut bien noter que dans ce cas, il n'est pas nécessaire d'imposer l'irréductibilité pour que tous les points soient de même nature.

On pose

$$\phi(t) = \widehat{\nu}(t) = \int_{\mathbb{R}^d} \mathrm{e}^{i<t,x>}\, d\nu(x) = \sum_{x \in \mathbb{Z}^d} \mathrm{e}^{i<t,x>} \nu(x),$$

on a donc

$$\nu(x) = (2\pi)^{-d} \int_{\mathbb{T}^d} \mathrm{e}^{-i<t,x>} \phi(t)\, dt, \quad \mathbb{T}^d = [-\pi, +\pi[^d.$$

On a alors $Q^n(0,0) = \mathbb{P}_0(X_n = 0) = \mathbb{Q}(S_n = 0)$. Mais Y_k a pour fonction caractéristique $\phi(t)$ et S_n a pour f.c. $\phi^n(t)$ d'où $Q^n(0,0) = \mathbb{Q}(S_n = 0) = (2\pi)^{-d} \int_{\mathbb{T}^d} \phi^n(t)\, dt$.
On en déduit, pour $0 < r < 1$,

$$U_r(0,0) := \sum_{n \geq 0} r^n \mathbb{P}_0(X_n = 0) = \sum_{n \geq 0} r^n (2\pi)^{-d} \int_{\mathbb{T}^d} \phi^n(t)\, dt$$

$$= (2\pi)^{-d} \int_{\mathbb{T}^d} \sum_{n \geq 0} r^n \phi^n(t)\, dt = (2\pi)^{-d} \int_{\mathbb{T}^d} \frac{1}{1 - r\phi(t}\, dt = (2\pi)^{-d} \int_{\mathbb{T}^d} \Re[\frac{1}{1 - r\phi(t)}]\, dt$$

à condition de justifier ce calcul. Mais, vu que $|\phi(t)| \leq 1$,

$$|\sum_{n=0}^{m} r^n \phi^n(t)| \leq \sum_{n=0}^{m} r^n |\phi^n(t)| \leq \frac{1}{1-r}$$

et on peut appliquer le th. de Lebesgue (on peut aussi majorer le reste).
Vu que (convergence monotone), lorsque $r \uparrow 1$,

$$U_r(0,0) = \sum_{n \geq 0} r^n Q^n(0,0) \uparrow \sum_{n \geq 0} Q^n(0,0) = U(0,0)$$

On a alors

Proposition 4.9.4. *Soit S_n une marche aléatoire sur $\mathbb{Z}^d$ de loi ν. Les points de $\mathbb{Z}^d$ sont récurents ssi*

$$\lim_{r \uparrow 1} \int_{\mathbb{T}^d} \Re[\frac{1}{1 - r\widehat{\nu}(t)}]\, dt = +\infty$$

Considérons maintenant la marche aléatoire élémentaire sur $\mathbb{Z}^d$. Soient

$$e_1 = (1, 0, \ldots, 0), \ldots, e_d = (0, \ldots, 0, 1) \quad \text{et} \quad \nu(e_i) = \nu(-e_i) = \frac{1}{2d}, \quad i = 1, \ldots, d$$

On a

$$\phi(t) = \frac{1}{d}\sum_{k=1}^{d} \cos t_k$$

Sur le complémentaire d'un voisinage de 0 de type $C_\alpha = \{||t|| \leq \alpha\}$, dans $\mathbb{T}^d$, la fonction $t \to (1 - r\phi(t))$ est bornée inférieurement uniformément en $r \leq 1$ et par conséquent la limite de l'intégrale est finie sur cet ensemble. Sur un tel voisinage, si α est suffisament petit, la fonction $r \to 1/(1-r\phi(t))$ est croissante positive et donc le théorème de limite monotone permet de se ramener à l'intégrabilité de $1/(1-\phi(t))$ sur C_α. Or $1/(1-\phi(t))$ est équivalente à $2d||t||^{-2}$ et $\int_{C_\alpha} \frac{1}{1-\phi(t)}\, dt = \infty$ ssi $d \leq 2$ puisque, en utilisant les coordonnées polaires dans $\mathbb{R}^d$ on obtient $\int_{C_\alpha} ||t||^{-2}\, dt \sim \int_0^\alpha r^{d-3}\, dr$. Donc la marche est récurrente si $d \leq 2$, transiente si $d \geq 3$. Il faut noter que, pour tout d, la mesure de comptage λ sur $\mathbb{Z}^d$ définie par, pour tout $x \in \mathbb{Z}^d$, $\lambda(x) = 1$, est une mesure invariante.

Chapitre 5

Martingales

5.1. Définition et premières propriétés

Soit $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$ un espace de probabilité filtré.

Définition 5.1.1. *Un processus réel adapté X_n est une martingale (resp. une surmartingale, resp. une sous-martingale) si, pour tout n, X_n est intégrable et*

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \text{ } p.s. \text{ } (resp. \ \leq, \ resp. \ \geq)$$

(Le cas positif et non nécessairement intégrable sera traité plus loin).

Donc si X_n est une surmartingale, $-X_n$ est une sous-martingale et réciproquement; X_n est une martingale ssi X_n est à la fois une surmartingale et une sous-martingale.

Z. La notion de martingale est relative à une filtration $\mathcal{F}_n$ et il serait plus correct de parler de $\mathcal{F}_n$-martingale; ce que l'on fera s'il y a la moindre ambiguïté. Si aucune filtration n'est précisée, on dit qu'un processus réel $X = (\Omega, \mathcal{F}, X_n, \mathbb{P})$ est une martingale si c'est une $\mathcal{F}_n^0$-martingale où $\mathcal{F}_n^0 = \sigma(X_0, \ldots, X_n)$.

Si X_n est une martingale, $\mathbb{E}(X_n)$ est constant, si X_n est une surmartingale, $\mathbb{E}(X_n)$ est décroissant, si X_n est une sous-martingale, $\mathbb{E}(X_n)$ est croissant.
Soit X_n est une surmartingale, on a pour $n > m$,

$$\mathbb{E}(X_n|\mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m) \leq \mathbb{E}(X_{n-1}|\mathcal{F}_m) \ldots \leq \mathbb{E}(X_{m+1}|\mathcal{F}_m) \leq X_m \text{ p.s.}$$

Il en résulte qu'un processus intégrable X_n est une surmartingale ssi

$$\text{pour tous } n > m, \text{ pour tout } A \in \mathcal{F}_m, \quad \int_A X_n \, d\mathbb{P} \leq \int_A X_m \, d\mathbb{P}. \tag{5.1}$$

Proposition 5.1.2. *Soient X_n une martingale (resp. une sous-martingale) et $f : \mathbb{R} \to \mathbb{R}$ convexe (resp. convexe croissante). On suppose $f(X_n)$ intégrable. Alors $f(X_n)$ est une sous-martingale.*

`Preuve:`
D'après l'inégalité de Jensen, $\mathbb{E}(f(X_{n+1})|\mathcal{F}_n) \geq f(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = f(X_n)$ p.s. si X_n est une martingale et $f(\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq f(X_n)$ p.s. si f est croissante et X_n une sous martingale puisque, dans ce cas, $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$ p.s. ∎

Ceci implique que, si X_n est une martingale, $|X_n|^p$, $p \geq 1$, est une sous-martingale si $X_n \in L^p$ et que, si X_n est une sous-martingale positive, X_n^p, $p > 1$, est une sous-martingale si $X_n \in L^p$.

Définition 5.1.3. *Soit $1 \leq p < \infty$. On dit que le processus X_n est borné dans L^p, si $\sup_n \mathbb{E}(|X_n|^p) < \infty$.*

Z. S'il existe $U \in L^p$ avec $|X_n| \leq U$, alors la suite X_n est bornée dans L^p, mais ces deux propriétés ne sont pas du tout équivalentes! (voir par contre (5.3.2)).

Lemme 5.1.4. *Une sous-martingale (resp. une surmartingale) est bornée dans L^1 ssi $\sup_n \mathbb{E}(X_n^+) < +\infty$, (resp. $\sup_n \mathbb{E}(X_n^-) < +\infty$).*

`Preuve:`
Soit X_n une sous-martingale, $\mathbb{E}(X_n) \geq \mathbb{E}(X_0)$ et $\mathbb{E}|X_n| = 2\mathbb{E}(X_n^+) - \mathbb{E}(X_n) \leq 2\mathbb{E}(X_n^+) - \mathbb{E}(X_0)$ ∎

Exemples

• Soit $X \in L^1$. On pose $X_n = \mathbb{E}(X|\mathcal{F}_n)$. On a

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_{n+1})|\mathcal{F}_n) = \mathbb{E}(X|\mathcal{F}_n) = X_n \text{ p.s.}$$

et X_n est une martingale.

• Soit $Y_0, Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r intégrables adaptées. On suppose que, pour tout $n \geq 1$, $\mathbb{E}(Y_n|\mathcal{F}_{n-1}) = 0$ (resp. $\mathbb{E}(Y_n|\mathcal{F}_{n-1}) \leq 0$, resp. $\mathbb{E}(Y_n|\mathcal{F}_{n-1}) \geq 0$), alors $X_n = \sum_{k=0}^n Y_k$ est une martingale (resp. une surmartingale, resp. une sous-martingale). (Vérification immédiate).
Réciproquement si X_n est une martingale (resp. une surmartingale, resp. une sous-martingale), posant $Y_0 = X_0$ et, pour $n \geq 1$, $Y_n = X_n - X_{n-1}$, on a $X_n = \sum_{k=0}^n Y_k$ avec pour tout $n \geq 1$, $\mathbb{E}(Y_n|\mathcal{F}_{n-1}) = 0$ (resp. $\mathbb{E}(Y_n|\mathcal{F}_{n-1}) \leq 0$, resp. $\mathbb{E}(Y_n|\mathcal{F}_{n-1}) \geq 0$).

• Soit $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes intégrables, avec $\mathbb{E}(Y_n) = \mu_n$. On pose $m_n = \sum_{k=1}^n \mu_k$ puis:

$$S_0 = 0,\ \ \mathcal{F}_0 = \{\Omega, \emptyset\} \text{ et pour } n \geq 1,\ S_n = Y_1 + \ldots + Y_n,\ \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n),$$

La variable aléatoire Y_n est indépendante de $\mathcal{F}_{n-1}$ et donc $\mathbb{E}(f(Y_n)|\mathcal{F}_{n-1}) = \mathbb{E}(f(Y_n))$ pour toute fonction f telle que $f(Y_n)$ soit intégrable. On en déduit que:

1. $Z_n = S_n - m_n$ est une martingale. En effet $Z_n - Z_{n-1} = Y_n - \mu_n$ et donc $\mathbb{E}((Z_n - Z_{n-1})|\mathcal{F}_{n-1}) = \mathbb{E}(Y_n) - \mu_n = 0$.

2. Si la suite Y_n est de carré sommable, on pose $\sigma_n^2 = \mathrm{Var}(Y_k)$ et $v_n^2 = \sum_{k=1}^n \sigma_k^2$. Alors $X_n = Z_n^2 - v_n^2$ est une martingale. En effet:

$$\begin{aligned} X_n - X_{n-1} &= 2Z_{n-1}(Y_n - \mu_n) + (Y_n - \mu_n)^2 - \sigma_n^2 \\ \mathbb{E}((X_n - X_{n-1})|\mathcal{F}_{n-1}) &= 2Z_{n-1}\mathbb{E}((Y_n - \mu_n)|\mathcal{F}_{n-1}) + \mathbb{E}((Y_n - \mu_n)^2|\mathcal{F}_{n-1}) - \sigma_n^2 \\ &= 2Z_{n-1}\mathbb{E}(Y_n - \mu_n) + \mathbb{E}((Y_n - \mu_n)^2) - \sigma_n^2 = 0 \end{aligned}$$

   On verra plus loin que cette propriété correspond à la décomposition de Doob de la sous-martingale Z_n^2.

• On se place comme ci dessus dans le cas de variables aléatoires indépendantes mais on suppose de plus qu'elles suivent toutes une loi normale, centrée, réduite. On pose $X_0 = 1$ puis, pour $n \geq 1$, $X_n = \exp(S_n - n/2)$. Alors X_n est une martingale (positive) puisque:

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_{n-1}\exp(Y_n - 1/2)|\mathcal{F}_{n-1}) = X_{n-1}\mathbb{E}(\exp(Y_n - 1/2)) = X_{n-1}$$

Cette martingale, souvent appelée "martingale exponentielle", est très utilisée en calcul stochastique, mais elle est aussi à la base de nombreux contre-exemples. En effet, d'après la loi forte des grands nombres, la suite S_n/n converge p.s. vers 0 et donc $X_n = \exp(n(S_n/n - 1/2))$ converge aussi p.s. vers 0 alors que son espérance est constante, égale à 1. On n'a donc pas convergence dans L^1, bien que cette martingale soit bornée dans L^1.

Décomposition de Doob d'une sous-martingale

On dit qu'un processus A_n est un processus croissant prévisible si $A_0 = 0$, $A_n \leq A_{n+1}$ et si A_{n+1} est $\mathcal{F}_n$-mesurable. On note $A_\infty = \lim \uparrow A_n$.

Proposition 5.1.5. *Toute sous-martingale X_n s'écrit, de façon unique, sous la forme $X_n = M_n + A_n$ avec M_n martingale et A_n processus croissant prévisible.*

`Preuve:`
On définit $A_0 = 0$, $A_{n+1} = A_n + \mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n)$, alors, par construction, A_n est un processus croissant prévisible intégrable et $M_n = X_n - A_n$ vérifie $\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0$ (puisque A_{n+1} est $\mathcal{F}_n$ mesurable). L'écriture est unique car si $X_n = M_n + A_n = M'_n + A'_n$, on a $A_0 = A'_0 = 0$ et $A'_{n+1} - A'_n = X_{n+1} - X_n - (M'_{n+1} - M'_n)$ d'où, en conditionnant par $\mathcal{F}_n$, $A'_{n+1} - A'_n = \mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) = A_{n+1} - A_n$ d'où $A'_n = A_n$ puis $M'_n = M_n$ ∎

Martingales arrêtées

Si X_n est un processus et T un temps d'arrêt, on définit le processus arrêté X^T par $X_n^T = X_{n\wedge T}$. On a alors

Proposition 5.1.6. *Soient X_n une surmartingale (resp. martingale) et T un temps d'arrêt, alors X^T est une surmartingale (resp. martingale).*

Preuve:
On a $|X_{n\wedge T}| \leq |X_0| + \ldots + |X_n|$ donc X^T est intégrable. On a alors, vu que $\{T > n\} = \{T \leq n\}^c \in \mathcal{F}_n$,

$$\mathbb{E}(X^T_{n+1}|\mathcal{F}_n) = \mathbb{E}(X_{(n+1)\wedge T}|\mathcal{F}_n) = \sum_{k=0}^{n} \mathbb{E}(X_k \mathbb{1}_{\{T=k\}}|\mathcal{F}_n) + \mathbb{E}(X_{n+1}\mathbb{1}_{\{T>n\}}|\mathcal{F}_n)$$

$$= \sum_{k=0}^{n} X_k \mathbb{1}_{\{T=k\}} + \mathbb{1}_{\{T>n\}}\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq \sum_{k=0}^{n} X_k \mathbb{1}_{\{T=k\}} + \mathbb{1}_{\{T>n\}} X_n = X_{n\wedge T} = X^T_n \;\blacksquare$$

5.2. Etude sur un intervalle de temps fini

Soit X_n une martingale. On a pour tout n, $\mathbb{E}(X_n) = \mathbb{E}(X_0)$. La question est de savoir si on a encore $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ pour T temps d'arrêt. La réponse est en général négative mais cette propriété est cependant exacte pour des temps d'arrêt à valeurs dans un intervalle borné de $\mathbb{N}$.

Le théorème d'arrêt borné

Définition 5.2.1. *Un temps d'arrêt T est dit borné s'il existe $M \in \mathbb{R}^+$ tel que, pour presque tout $\omega \in \Omega$, on ait $T(\omega) \leq M$.*

Théorème 5.2.2. *Soient X_n une surmartingale (resp. martingale) et $T_1 \leq T_2$ deux temps d'arrêt bornés, alors X_{T_1} et X_{T_2} sont intégrables et:*

$$\mathbb{E}(X_{T_2}|\,\mathcal{F}_{T_1}) \leq X_{T_1} \text{ p.s. (resp. } =). \tag{5.2}$$

Preuve:
On suppose $0 \leq T_1 \leq T_2 \leq m \in \mathbb{N}$. On a $|X_{T_i}| \leq |X_0| + \ldots + |X_m| \in L^1$ pour $i = 1, 2$. Soit alors $A \in \mathcal{F}_{T_1}$, vu la prop.5.1.6,

$$\int_A X_{T_2}\,dP = \sum_{k=0}^{m} \int_{A\cap\{T_1=k\}} X_{T_2\wedge m}\,dP \leq \sum_{k=0}^{m} \int_{A\cap\{T_1=k\}} X_{T_2\wedge k}\,dP = \int_A X_{T_1}\,dP$$

puisque $A \cap \{T_1 = k\} \in \mathcal{F}_k$ $\blacksquare$

Si X_n est une surmartingale et T un temps d'arrêt tel que $\mathbb{P}(T < +\infty) = 1$ mais pas borné (bien comprendre la différence), on sait que, pour tout n, $\mathbb{E}(X_{T\wedge n}) \leq \mathbb{E}(X_0)$ (puisque $X_{T\wedge n}$ est une surmartingale). Alors, T étant p.s. fini, $X_{T\wedge n} \to_n X_T$ p.s. et tout le problème est de savoir si $\mathbb{E}(X_{T\wedge n}) \to \mathbb{E}(X_T)$. Ce n'est pas toujours vrai mais on a:

Corollaire 5.2.3. *Soient X_n une surmartingale (resp. martingale) et T un temps d'arrêt. On suppose $\mathbb{P}(T < +\infty) = 1$, $\mathbb{E}|X_T| < +\infty$ et $\mathbb{E}(X_n \mathbb{1}_{\{T>n\}}) \to_n 0$. Alors $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ (resp. $=$).*

Preuve:
On a $\mathbb{E}(X_{T\wedge n}) \leq \mathbb{E}(X_0)$ et $\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_T \mathbb{1}_{\{T\leq n\}}) + \mathbb{E}(X_n \mathbb{1}_{\{T>n\}})$. Vu que $\mathbb{P}(T < +\infty) = 1$, $X_T \mathbb{1}_{\{T\leq n\}} \to X_T$ p.s. et $|X_T \mathbb{1}_{\{T\leq n\}}| \leq |X_T| \in L^1$ d'où $\mathbb{E}(X_T \mathbb{1}_{\{T\leq n\}}) \to_n \mathbb{E}(X_T)$. Donc, vu la dernière hypothèse, $\mathbb{E}(X_T) = \lim_n \mathbb{E}(X_{T\wedge n}) \leq E(X_0)$ ∎

Inégalités maximales

On appelle inégalités maximales des inégalités relatives à $\sup_n X_n$ ou $\sup_n |X_n|$.

Proposition 5.2.4. *Pour un processus X_n, on pose $X_n^* = \sup_{0\leq k\leq n} X_k$.*

1. *Soit X_n une surmartingale. On a, pour tout $a > 0$ et tout n,*

$$a\mathbb{P}(X_n^* > a) \leq \mathbb{E}(X_0) - \int_{(X_n^* \leq a)} X_n \, d\mathbb{P} \leq \mathbb{E}(X_0) + \mathbb{E}(X_n^-)$$

2. *Soit X_n une sous-martingale. On a, pour tout $a > 0$ et tout n,*

$$a\mathbb{P}(X_n^* > a) \leq \int_{(X_n^* > a)} X_n \, d\mathbb{P} \leq \mathbb{E}(X_n^+)$$

Preuve:
Soit $T = \inf(k \geq 0,\ X_k > a) \wedge n$. Le temps d'arrêt T est borné par n et de plus on a les inclusions $(X_n^* \leq a) \subset (T = n)$ et $(X_n^* > a) \subset (X_T > a)$. On en déduit

$$\mathbb{E}(X_T) = \int_{(X_n^* > a)} X_T \, d\mathbb{P} + \int_{(X_n^* \leq a)} X_T \, d\mathbb{P} \geq a\mathbb{P}(X_n^* > a) + \int_{(X_n^* \leq a)} X_n \, d\mathbb{P}$$

D'après le théorème d'arrêt borné,
• lorsque X est une surmartingale on a $\mathbb{E}(X_0) \geq \mathbb{E}(X_T)$ et donc:

$$\mathbb{E}(X_0) \geq a\mathbb{P}(X_n^* > a) + \int_{(X_n^* \leq a)} X_n \, d\mathbb{P}$$

• lorsque X est une sous martingale on a $\mathbb{E}(X_n) \geq \mathbb{E}(X_T)$ et donc:

$$\mathbb{E}(X_n) \geq a\mathbb{P}(X_n^* > a) + \int_{(X_n^* \leq a)} X_n \, d\mathbb{P}$$

Ceci fournit les premiers membres des inégalités ci dessus. Pour obtenir le second membre dans chaque inégalité, il suffit de constater que pour toute variable aléatoire intégrable X et tout $A \in \mathcal{F}$ on a:

$$-\mathbb{E}(X^-) \leq -\int_A X^- \, d\mathbb{P} \leq \int_A X \, d\mathbb{P} \leq \int_A X^+ \, d\mathbb{P} \leq \mathbb{E}(X^+)$$

5.3. Martingales dans L^2

Nous commençons par étudier la structure des martingales dans L^2.

Proposition 5.3.1. *Soit X_n une martingale dans L^2. On pose $\Delta_0 = X_0$ et pour $n \geq 1$, $\Delta_n = X_n - X_{n-1}$. Alors $X_n = \Delta_0 + \Delta_1 + \ldots + \Delta_n$ et les variables aléatoires Δ_k, $k \geq 0$ forment un système orthogonal dans L^2.*

`Preuve:`
Soit $m > n$. On a $\mathbb{E}(\Delta_m|\mathcal{F}_n) = 0$, d'où:

$$\mathbb{E}(\Delta_m \Delta_n) = \mathbb{E}(\mathbb{E}(\Delta_m \Delta_n|\mathcal{F}_n)) = \mathbb{E}(\Delta_n \mathbb{E}(\Delta_m|\mathcal{F}_n)) = 0 \quad \blacksquare$$

On peut remarquer que, pour $n \geq 1$, on a $\mathbb{E}(\Delta_n) = 0$. De plus le théorème de Pythagore permet d'affirmer que $\mathbb{E}(X_n^2) = \sum_{k=0}^n \mathbb{E}(\Delta_k^2)$. La martingale X_n est donc bornée dans L^2 (c'est à dire $\sup_n \mathbb{E}(X_n^2) < +\infty$) si et seulement si la série $\sum \mathbb{E}(\Delta_n^2)$ est convergente.

Z. Il est faux de dire que les sommes partielles d'un système orthogonal centré forment une martingale (contrairement à un système indépendant).

On va maintenant étudier la classe des martingales bornées dans L^2. Le résultat suivant n'est pas vraiment nécessaire pour la suite, mais il explique la différence de comportement des martingales bornées dans L^2 ou bien seulement bornées dans L^1. En effet, ce résultat est généralement faux dans L^1 (voir par exemple le cas de la martingale exponentielle citée en exemple dans la section précédente).

Proposition 5.3.2. *Soit X_n une martingale dans L^2. Alors $\mathbb{E}(\sup_n |X_n|^2) \leq 4 \sup_n \mathbb{E}(X_n^2)$. Par conséquent, si la martingale X_n est bornée dans L^2, il existe une variable aléatoire $U \in L^2$ avec $|X_n| \leq U$.*

`Preuve:`
La suite $|X_n|$ est une sous-martingale. Par conséquent, en posant $Y_n = \max_{k \leq n} |X_k|$ on a, d'après la proposition 5.2.4, la majoration $a\mathbb{P}(Y_n > a) \leq \mathbb{E}(|X_n| \mathbb{1}_{\{Y_n > a\}})$ pour $a > 0$. En intégrant celle-ci sur $a \in]0, \infty[$, on obtient:

$$\int_0^{+\infty} aE(\mathbb{1}_{\{Y_n > a\}})\, da \leq \int_0^{+\infty} \mathbb{E}(|X_n| \mathbb{1}_{\{Y_n > a\}})\, da$$

soit encore, en utilisant le théorème de Fubini et l'inégalité de Schwarz:

$$\mathbb{E}(Y_n^2)/2 \leq \mathbb{E}(|X_n| Y_n) \leq ||X_n||_2 ||Y_n||_2$$

D'où finalement $\mathbb{E}(Y_n^2) \leq 4\mathbb{E}(X_n^2)$. La suite Y_n étant croissante, le théorème de limite monotone permet d'obtenir la majoration cherchée puisque $\sup_n Y_n = \sup_n |X_n|$. Par conséquent, si X_n est bornée dans L^2, on peut choisir $U = \sup |X_n|$ $\blacksquare$

On rappelle un critère de convergence p.s. établi dans le second chapitre (voir 2.5.8).

Lemme 5.3.3. *Une suite* X_n *de v.a.r. converge p.s. ssi, pour tout* $\epsilon > 0$,

$$\lim_n \mathbb{P}(\sup_{p\geq 0} |X_{p+n} - X_n| > \epsilon) = 0$$

Théorème 5.3.4. *Soit* X_n *une martingale bornée dans* L^2*. Alors* X_n *converge p.s. et dans* L^2 *vers une variable aléatoire* X_∞ *telle que* $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$*.*

`Preuve:`
D'après la proposition 5.3.1 on a pour $m > n$:

$$\mathbb{E}((X_m - X_n)^2) = \mathbb{E}((\sum_{k=n+1}^{m} \Delta_k)^2) = \sum_{k=n+1}^{m} \mathbb{E}(\Delta_k)^2$$

Il suffit donc d'utiliser la propriété de Cauchy de la suite convergente $\sum_n \mathbb{E}(\Delta_n^2)$ pour obtenir que X_n est de Cauchy dans L^2 et donc convergente dans L^2 vers une variable aléatoire X_∞. Pour obtenir la convergence p.s. on remarque que pour n fixé, la suite $Z_p = X_{n+p} - X_n$ est une martingale de carré sommable, et par conséquent, on peut appliquer l'inégalité maximale à la sous martingale Z_p^2 soit:

$$\epsilon^2 \mathbb{P}(\sup_{0\leq p\leq m} |X_{n+p} - X_n| > \epsilon) \leq \mathbb{E}(Z_m^2) = \sum_{k=n+1}^{m+n} \mathbb{E}(\Delta_k^2)$$

En faisant tendre m vers l'infini on obtient:

$$\epsilon^2 \mathbb{P}(\sup_{p\geq 0} |X_{n+p} - X_n| > \epsilon) \leq \sum_{k=n+1}^{\infty} \mathbb{E}(\Delta_k^2)$$

et il ne reste plus quà utiliser le fait que le second membre est le reste d'une série convergente pour avoir la convergence p.s.(nécessairement vers la variable aléatoire X_∞ obtenue préceédemment). Pour obtenir la dernière affirmation on utilise le fait que la convergence L^2 implique la convergence L^1 et que pour tout $A \in \mathcal{F}_n$ et tout k, $\int_A X_n \, d\mathbb{P} = \int_A X_{n+k} \, d\mathbb{P}$. On obtient pour tout k:

$$|\int_A X_n \, d\mathbb{P} - \int_A X_\infty \, d\mathbb{P}| \leq \int_A |X_{n+k} - X_\infty| \, d\mathbb{P} \leq ||X_{n+k} - X_\infty||_1 \to_{k\to\infty} 0$$

ce qui montre que $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$ ∎

On peut remarquer que d'après le théorème de Lebesgue dans L^2 (2.6.5), la convergence p.s. jointe à la proposition 5.3.2 impliquait la convergence dans L^2.

Corollaire 5.3.5. *Soit* $X \in L^2$*. Alors la martingale* $X_n = \mathbb{E}(X|\mathcal{F}_n)$ *est bornée dans* L^2 *et sa limite* X_∞ *vérifie* $X_\infty = \mathbb{E}(X|\mathcal{F}_\infty)$ *(et donc* $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$*).*

`Preuve:`
D'après l'inégalité de Jensen, on a $(\mathbb{E}(X|\mathcal{F}_n))^2 \leq \mathbb{E}(X^2|\mathcal{F}_n)$ et en prenant l'espérance $\mathbb{E}(X_n^2) \leq \mathbb{E}(X^2)$, ce qui prouve que X_n est bornée dans L^2. Pour tout $A \in \mathcal{F}_n$ on a donc d'après le théorème 5.3.4 l'égalité $\int_A X_n \, d\mathbb{P} = \int_A X \, d\mathbb{P} = \int_A X_\infty \, d\mathbb{P}$. Ceci implique que les deux mesures $\mu(A) = \int_A X \, d\mathbb{P}$ et $\nu(A) = \int_A X_\infty \, d\mathbb{P}$ coïncident sur l'algèbre de Boole $\cup_n \mathcal{F}_n$ et donc sur la σ-algèbre engendrée soit $\mathcal{F}_\infty$. La variable aléatoire X_∞ étant $\mathcal{F}_\infty$ mesurable, on a la propriété demandée ∎

Décomposition de Doob de X_n^2

Définition 5.3.6. *Soit X_n une martingale de carré intégrable. On note $X_n^2 = M_n + A_n$, la décomposition de Doob de la sous-martingale X_n^2 fournie par la proposition 5.1.5.*

On a alors
$$A_{n+1} - A_n = \mathbb{E}(X_{n+1}^2 - X_n^2|\mathcal{F}_n) = \mathbb{E}((X_{n+1} - X_n)^2|\mathcal{F}_n) \tag{5.3}$$
puisque, pour une martingale de L^2,
$$\mathbb{E}(X_n X_{n+1}|\mathcal{F}_n) = X_n \mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n^2 \text{ p.s.}$$

La preuve du théorème suivant exige l'utilisation d'un lemme, dit "lemme de Kronecker" que nous reproduisons ci-dessous:

Lemme 5.3.7. *On considère une suite reélle x_n et des coefficients α_n strictement positifs, formant une suite croissante qui tend vers $+\infty$. Si la série $\sum \frac{x_n}{\alpha_n}$ converge, alors la suite $\frac{1}{\alpha_n} \sum_{k=1}^n x_k$ tend vers 0.*

`Preuve:`
La suite $v_n = \alpha_n - \alpha_{n-1}$ est positive et la série $\sum_n v_n$ est divergente. D'autre part, la suite $z_n = \sum_{k=1}^n \frac{x_k}{\alpha_k}$ est convergente et en posant $\alpha_0 = z_0 = 0$, l'on a:
$$\sum_{k=1}^n x_k = \sum_{k=1}^n \alpha_k (z_k - z_{k-1}) = \sum_{k=1}^n v_k (z_n - z_{k-1})$$
Par conséquent pour $p < n$:
$$|\sum_{k=1}^n x_k| \leq |\sum_{k=1}^p v_k (z_n - z_{k-1})| + (\sum_{k=p+1}^n v_k) \sup_{p \leq k \leq n-1} |z_n - z_k|$$
Pour tout $\epsilon > 0$, on a $\sup_{p \leq k \leq n-1} |z_n - z_k| \leq \epsilon$ pour n et p assez grands. On obtient le résultat en divisant les deux membres par α_n et on fait tendre n puis p vers l'infini ∎

On pose $A_\infty = \lim \uparrow_n A_n$. La variable aléatoire A_∞ donne des informations sur le comportement asymptotique de X_n. D'abord si $\mathbb{E}(A_\infty) < +\infty$, on a $\mathbb{E}(X_n^2) = \mathbb{E}(X_0^2) + \mathbb{E}(A_n)$ d'où $\mathbb{E}(X_n^2) \leq \mathbb{E}(X_0^2) + \mathbb{E}(A_\infty)$ et X_n est bornée dans L^2 donc (th.5.5.2) X_n converge p.s. et dans L^2 vers X_∞. Plus généralement:

Théorème 5.3.8. *Soit X_n une martingale de carré intégrable. Alors*

1. *sur l'ensemble $\{A_\infty < +\infty\}$, la suite X_n converge p.s. vers une limite finie.*
2. *sur l'ensemble $\{A_\infty = +\infty\}$, la suite $X_n/A_n \to_n 0$ p.s.*

`Preuve:`
(1) Pour un entier $N \geq 1$, on pose $T_N = \inf(n \geq 0,\ A_{n+1} > N)$. T_N est un temps d'arrêt car $\{T_N = n\} = \{A_1 \leq N, \ldots, A_n \leq N, A_{n+1} > N\} \in \mathcal{F}_n$ et de plus on a $A_{n\wedge T_N} \leq N$, d'où:

$$\mathbb{E}(X_{n\wedge T_N}^2) = \mathbb{E}(X_0^2) + \mathbb{E}(A_{n\wedge T_N}) \leq \mathbb{E}(X_0^2) + N$$

La martingale $X_{n\wedge T_N}$ est par conséquent bornée dans L^2 et il existe donc un ensemble Ω_N avec $\mathbb{P}(\Omega_N) = 1$ et tel que pour tout $\omega \in \Omega_N$, la suite $X_{n\wedge T_N(\omega)}(\omega)$ converge vers une limite finie. On a d'une part $\mathbb{P}(\cap_N \Omega_N) = 1$ et d'autre part

$$\{A_\infty < +\infty\} = \cup_N \{A_\infty \leq N\} = \cup_N \{T_N = +\infty\}$$

On en conclut que X_n converge p.s. sur l'ensemble $\{A_\infty < +\infty\}$, car si on choisit $\omega \in (A_\infty < +\infty) \cap (\cap_N \Omega_N)$, il existe N tel que $T_N(\omega) = \infty$. Or, pour cette valeur de N on a $X_{n\wedge T_N(\omega)}(\omega) = X_n(\omega)$ et de plus $\omega \in \Omega_N$.

(2) On pose $Z_0 = 0$ et pour $n \geq 1$, $Z_n = \sum_{k=1}^n V_k$ avec $V_n = \dfrac{X_n - X_{n-1}}{1 + A_n}$. On a $\mathbb{E}(V_n^2) \leq \mathbb{E}((X_n - X_{n-1})^2)$ et par conséquent V_n et donc Z_n sont de carré intégrable.

$$\begin{aligned}
\mathbb{E}((Z_{n+1} - Z_n)|\mathcal{F}_n) &= \mathbb{E}(V_{n+1}|\mathcal{F}_n) = \frac{1}{1 + A_{n+1}} \mathbb{E}((X_{n+1} - X_n)|\mathcal{F}_n) = 0 \\
\mathbb{E}((Z_{n+1} - Z_n)^2|\mathcal{F}_n) &= \mathbb{E}(V_{n+1}^2|\mathcal{F}_n) = \frac{1}{(1 + A_{n+1})^2} \mathbb{E}((X_{n+1} - X_n)^2|\mathcal{F}_n) = \frac{A_{n+1} - A_n}{(1 + A_{n+1})^2}
\end{aligned}$$

La première ligne montre que Z_n est une martingale, la seconde que le processus croissant B_n associé à la décomposition de Doob de Z_n^2 est défini par:

$$B_n = \sum_{k=1}^n \frac{A_k - A_{k-1}}{(1 + A_k)^2} \uparrow \sum_{k=1}^\infty \frac{A_k - A_{k-1}}{(1 + A_k)^2} = B_\infty$$

Or pour toute suite croissante a_k de réels positifs, on a:

$$\sum_{k=1}^\infty \frac{a_k - a_{k-1}}{(1 + a_k)^2} \leq \sum_{k=1}^\infty \int_{a_{k-1}}^{a_k} \frac{dt}{(1 + t)^2} \leq \int_0^\infty \frac{dt}{(1 + t)^2} = 1$$

On a donc $B_\infty \leq 1$ et la martingale Z_n est bornée dans L^2. Il s'en suit que la série $\sum V_n$ converge p.s. On applique alors le lemme de Kronecker avec $\alpha_n = 1 + A_n$ et $x_k = X_k - X_{k-1}$. On a que, sur l'ensemble $\{A_\infty = +\infty\}$, la suite $\dfrac{X_n}{1 + A_n} \to 0$ p.s. ce qui implique que $\dfrac{X_n}{A_n} \to 0$ p.s. ∎

5.4. Martingales dans L^1

La question est de savoir ce qui subsiste de la section précédente, lorsque l'on suppose seulement que la martingale X_n est bornée dans L^1. A priori, seule subsiste la convergence p.s., et si l'on veut obtenir la convergence L^1 l'on va devoir imposer une propriété supplémentaire, dite de "régularité".

Le résultat essentiel est:

Théorème 5.4.1. *Soit X_n une sous-martingale (resp. une surmartingale) bornée dans L^1, alors X_n converge p.s. vers une v.a. intégrable X_∞.*

La preuve de ce théorème va résulter d'une succession de lemmes.

Lemme 5.4.2. *Soit X_n une sous martingale telle qu'il existe une constante $K < \infty$ avec $|X_n| \leq K$. Alors la suite X_n converge p.s.*

`Preuve:`
Soit $X_n = M_n + A_n$ la décomposition de Doob (prop.5.1.5) de X_n. On a $\mathbb{E}(A_n) = \mathbb{E}(X_n) - \mathbb{E}(M_0) = \mathbb{E}(X_n) - \mathbb{E}(X_0) \leq 2K$ et donc $\mathbb{E}(A_\infty) < +\infty$ d'où $A_\infty < +\infty$ p.s. Comme dans 5.3.8, pour un entier $N > 0$, on pose $T_N = \inf(n \geq 0\,,\, A_{n+1} > N)$; T_N est un temps d'arrêt, $A_{T_N \wedge n} \leq N$ d'où:

$$|M_{n\wedge T_N}| = |X_{n\wedge T_N} - A_{n\wedge T_N}| \leq K + N$$

La martingale $M_{n\wedge T_N}$ est par conséquent bornée dans L^2 et il existe donc un ensemble Ω_N avec $\mathbb{P}(\Omega_N) = 1$ et tel que pour tout $\omega \in \Omega_N$, la suite $M_{n\wedge T_N(\omega)}(\omega)$ converge vers une limite finie. On a d'une part $\mathbb{P}(\cap_N \Omega_N) = 1$ et d'autre part

$$\{A_\infty < +\infty\} = \cup_N \{A_\infty \leq N\} = \cup_N \{T_N = +\infty\}$$

On en conclut que M_n converge p.s. sur l'ensemble $\{A_\infty < +\infty\}$, car si on choisit $\omega \in (A_\infty < +\infty) \cap (\cap_N \Omega_N)$, il existe N tel que $T_N(\omega) = \infty$. Or, pour cette valeur de N on a $M_{n\wedge T_N(\omega)}(\omega) = M_n(\omega)$ et de plus $\omega \in \Omega_N$. On en déduit la convergence p.s. de de X_n puisque la suite A_n converge p.s. ∎

Lemme 5.4.3. *Soit X_n une surmartingale positive. Alors X_n converge p.s. vers une variable aléatoire intégrable.*

`Preuve:`
La suite $Y_n = \mathrm{e}^{-X_n}$ est une sous-martingale vérifiant $0 \leq Y_n \leq 1$ (prop.5.1.2). Vu le lemme 5.4.2, Y_n converge p.s. et aussi X_n dans $\overline{\mathbb{R}}^+$. Mais $\mathbb{E}(X_\infty) = \mathbb{E}(\underline{\lim}\, X_n) \leq \underline{\lim}\, \mathbb{E}(X_n) \leq \mathbb{E}(X_0) < +\infty$ et $X_\infty < +\infty$ p.s. ∎

Lemme 5.4.4. *Soit X_n une sous-martingale bornée dans L^1, alors on a $X_n = Y_n - Z_n$ avec Y_n martingale positive et Z_n surmartingale positive.*

`Preuve:`
X_n^+ est une sous-martingale (prop.5.1.2) et donc (prop.5.1.5) $X_n^+ = M_n + A_n$. On a $\mathbb{E}(X_n^+) = \mathbb{E}(M_n) + \mathbb{E}(A_n) = \mathbb{E}(M_0) + \mathbb{E}(A_n)$ d'où $\mathbb{E}(A_n) \leq \mathbb{E}(X_n^+) + \mathbb{E}|M_0|$ et $\mathbb{E}(A_\infty) = \lim \uparrow \mathbb{E}(A_n) \leq \sup_n \mathbb{E}(X_n^+) + \mathbb{E}|M_0| < +\infty$. Soit $Y_n = M_n + \mathbb{E}(A_\infty|\mathcal{F}_n)$, Y_n est une martingale et $Y_n \geq M_n + \mathbb{E}(A_n|\mathcal{F}_n) = M_n + A_n = X_n^+ \geq X_n$ donc $Y_n \geq 0$ et $Z_n = Y_n - X_n$ est une surmartingale positive ∎

`Preuve du théorème 5.4.1:`
Supposons que X_n soit une sous-martingale bornée dans L^1. Appliquant le lemme 5.4.4, on obtient que X_n s'écrit comme la différence d'une martingale positive et d'une surmartingale positive, elle converge donc p.s. puisque chacun des deux termes a une limite intégrable et donc finie p.s. Si on note X_∞ sa limite on a $\mathbb{E}|X_\infty| = \mathbb{E}(\underline{\lim}\, |X_n|) \leq \underline{\lim}\, \mathbb{E}|X_n| \leq \sup_n \mathbb{E}|X_n| < +\infty$. ∎

Si X_n est une martingale bornée dans L^1, on a $X_n \to X_\infty$ p.s. et $X_\infty \in L^1$ mais on n'a pas toujours $X_n \to X_\infty$ dans L^1. Ceci est déjà patent dans le cas d'une martingale exponentielle mais on peut aussi considérer l'exemple suivant: Soient $\Omega = [0,1]$, $\mathcal{F}_n = \sigma([k2^{-n}, (k+1)2^{-n}[,\ k = 0, 1, \ldots, 2^n - 1)$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P}$ = mesure de Lebesgue et $X_n = 2^n \mathbb{1}_{\{[0,2^{-n}[\}}$. Alors X_n est une martingale car, pour tout $A \in \mathcal{F}_n$, $\int_A X_{n+1}\, d\mathbb{P} = \int_A X_n\, d\mathbb{P}$ (il suffit de le vérifier pour $A = [0, 2^{-n}[$). On a $\mathbb{E}(X_n) = \mathbb{E}|X_n| = 1$, $X_n \to 0$ p.s. mais $\mathbb{E}|X_n - 0| = 1$ et X_n ne converge pas dans L^1.

Martingales régulières

On étudie le problème de la convergence des martingales dans L^1. Rappelons que, si $X_n \to X_\infty$ dans L^1, alors, pour tout $A \in \mathcal{F}$, $\int_A X_n\, d\mathbb{P} \to_n \int_A X_\infty\, d\mathbb{P}$ car

$$|\int_A X_n\, d\mathbb{P} - \int_A X_\infty\, d\mathbb{P}| \leq \int_A |X_n - X_\infty|\, d\mathbb{P} \leq ||X_n - X_\infty||_1 \to_{k\to\infty} 0$$

On sait que si $X \in L^1$, $X_n = \mathbb{E}(X|\mathcal{F}_n)$ est une martingale.

Définition 5.4.5. *Une martingale de la forme $X_n = \mathbb{E}(X|\mathcal{F}_n)$, $X \in L^1$, s'appelle une martingale régulière.*

Théorème 5.4.6. *Soit X_n une martingale. Les propriétés suivantes sont équivalentes:*

1. *X_n est régulière.*
2. *X_n est uniformément intégrable*
3. *X_n converge dans L^1 vers X_∞. Dans ce cas on a nécessairement $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$.*

`Preuve:`
• 1. ⇒ 2. C'est une conséquence de 3.2.14

• 2. ⇒ 3. Toute famille unifomément intégrable est bornée dans L^1 d'après 2.6.2. La martingale X_n converge donc p.s. vers X_∞. Le théorème 2.6.5 permet d'affirmer que l'on a aussi convergence dans L^1. La dernière affirmation est prouvée ci après.

• 3. $\Rightarrow$ 1. Par hypothèse on a, pour tout $A \in \mathcal{F}_n$ et tout k, $\int_A X_n \, d\mathbb{P} = \int_A X_{n+k} \, d\mathbb{P}$ et, vu que $X_m \to_m X_\infty$ dans L^1, $\int_A X_{n+k} \, d\mathbb{P} \to_k \int_A X_\infty \, d\mathbb{P}$ d'où $\int_A X_n \, d\mathbb{P} = \int_A X_\infty \, d\mathbb{P}$ ce qui montre que $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$ ∎

Remarque 5.4.7. *On considère une martingale régulière de la forme $X_n = \mathbb{E}(X|\mathcal{F}_n)$, $X \in L^1$. Alors le théorème ci dessus montre que X_n converge p.s. et dans L^1 vers une v.a. X_∞. En fait on a $X_\infty = \mathbb{E}(X|\mathcal{F}_\infty)$ où $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, \, n \geq 0)$.*

`Preuve:`
Puisque X_∞ est $\mathcal{F}_\infty$ mesurable, il suffit d'établir que $\int_A X_\infty \, d\mathbb{P} = \int_A X \, d\mathbb{P}$ pour tout ensemble $A \in \mathcal{F}_\infty$. Or ceci est vrai lorsque $A \in \mathcal{F}_n$ puisqu'alors les deux intégrales valent $\int_A X_n \, d\mathbb{P}$. Il suffit alors d'utiliser de nouveau l'unicité du prolongement d'une mesure de l'algèbre à la tribu engendrée comme dans la fin de la preuve de 5.3.5. ∎

Théorème d'arrêt

Proposition 5.4.8. *Soit X une v.a. intégrable ou positive. Alors p.s.*

$$\mathbb{E}(X|\mathcal{F}_T)\mathbb{1}_{\{T=n\}} = \mathbb{E}(X|\mathcal{F}_n)\mathbb{1}_{\{T=n\}}, \; n \in \bar{\mathbb{N}}.$$

`Preuve:`
Supposons $X \geq 0$. Il est facile de constater que le produit d'une variable aléatoire $\mathcal{F}_T$ mesurable par $\mathbb{1}_{\{T=n\}}$ est en fait $\mathcal{F}_n$ mesurable. Les deux membres de l'égalité à prouver sont donc $\mathcal{F}_n$ mesurables. Pour prouver qu'ils sont égaux, il suffit de prouver qu'ils ont la même intégrale sur tout ensemble $A \in \mathcal{F}_n$. Or:

$$\int_{A \cap (T=n)} \mathbb{E}(X|\mathcal{F}_n) \, d\mathbb{P} = \int_{A \cap (T=n)} X \, d\mathbb{P} = \int_{A \cap (T=n)} \mathbb{E}(X|\mathcal{F}_T) \, d\mathbb{P}$$

la dernière égalité provenant du fait que l'ensemble $(A \cap (T = n))$ est en fait $\mathcal{F}_T$ mesurable ∎

Soient X_n une martingale régulière et T un temps d'arrêt. Alors $X_\infty := \lim_n X_n$ existe p.s. et $(X_n)_{n \in \bar{\mathbb{N}}}$ est une martingale. La v.a. X_T est définie sans ambiguïté par $X_T = X_n$ sur $\{T = n\}$, $n \in \bar{\mathbb{N}}$. On a alors

Théorème 5.4.9. *Soit $X_n = \mathbb{E}(X \,|\mathcal{F}_n)$, $X \in L^1$, une martingale régulière. Alors*

1. *si T est un temps d'arrêt, $X_T = \mathbb{E}(X \,|\mathcal{F}_T)$ p.s. et $X_T \in L^1$,*
2. *si $T_1 \leq T_2$ sont des temps d'arrêt, $\mathbb{E}(X_{T_2}|\mathcal{F}_{T_1}) = X_{T_1}$ p.s.*

`Preuve:`
• D'après la proposition 5.4.8 $X_T = \mathbb{E}(X \,|\mathcal{F}_T)$ p.s. et donc $X_T \in L^1$.

• On a $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$ et donc

$$\mathbb{E}(X_{T_2}|\mathcal{F}_{T_1}) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_{T_2})|\mathcal{F}_{T_1}) = \mathbb{E}(X|\mathcal{F}_{T_1}) = X_{T_1} \text{ p.s.} \quad ∎$$

Il est très facile de constater que ce résultat est faux sans hypothèse de régularité sur la martingale X_n. Prenons, par exemple, pour X_n la marche aléatoire centrée de pas ± 1 sur $\mathbb{Z}$ et pour T le temps d'entrée dans un point $a \neq 0$. Alors $\mathbb{P}(T < \infty) = 1$ et $X_T = a$. Il s'en suit que $\mathbb{E}(X_T) = a \neq X_0 = 0$

5.5. Martingales positives généralisées

Définition 5.5.1. *La définition 5.1.1 a encore un sens si on remplace "X_n intégrable" par "X_n positive" c'est à dire à valeurs dans $\overline{\mathbb{R}}^+$. On parlera alors de martingale (resp. surmartingale, resp. sous-martingale) positive généralisée.*

Surmartingales positives généralisées

Théorème 5.5.2. *Soit X_n une surmartingale positive généralisée, alors*

1. *X_n converge p.s. (dans $\overline{\mathbb{R}}^+$) vers une v.a. X_∞ et $\mathbb{E}(X_\infty | \mathcal{F}_n) \leq X_n$.*
2. *Si T est un temps d'arrêt, $\mathbb{E}(X_\infty | \mathcal{F}_T) \leq X_T$*
3. *Pour tous temps d'arrêt tels que $T_1 \leq T_2$, $\mathbb{E}(X_{T_2} | \mathcal{F}_{T_1}) \leq X_{T_1}$ p.s.*

`Preuve:`
1• Il suffit de reprendre la preuve du lemme 5.4.3 à l'exception de l'assertion relative à l'intégrabilité de la limite. D'après le lemme de Fatou, $\mathbb{E}(X_\infty | \mathcal{F}_n) = \mathbb{E}(\lim_p X_{p+n} | \mathcal{F}_n) \leq \underline{\lim}_p \mathbb{E}(X_{n+p} | \mathcal{F}_n) \leq X_n$ p.s.

2• Il suffit de prouver que pour $n \in \bar{\mathbb{N}}$, on a $\mathbb{E}(X_\infty | \mathcal{F}_T) \leq X_T$ p.s. sur $\{T = n\}$, mais ceci résulte de la prop.5.4.8 et de l'inégalité ci dessus.

3• Enfin $Z_n = X_{n \wedge T_2}$ est une surmartingale généralisée positive qui converge p.s. vers $Z_\infty = X_{T_2}$ et donc $\mathbb{E}(Z_\infty | \mathcal{F}_{T_1}) \leq Z_{T_1}$. Or $Z_{T_1} = X_{T_1 \wedge T_2} = X_{T_1}$ ∎

Corollaire 5.5.3. *Soient X_n une surmartingale positive généralisée et T un temps d'arrêt, alors $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$.*

Martingales positives généralisées

Z. On peut bien sûr appliquer les résultats du théorème (5.5.2) concernant les surmartingales positives généralisées aux martingales positives généralisées, mais il faut se garder de transformer les inégalités en égalités! Par exemple, l'on n'a pas, en général, l'égalité $\mathbb{E}(X_\infty | \mathcal{F}_n) = X_n$ mais seulement $\mathbb{E}(X_\infty | \mathcal{F}_n) \leq X_n$, et de même pour les temps d'arrêt.

Par contre, ces égalités deviennent vraies pour une forme particulière de martingale positive généralisée, proche de la notion de martingale régulière.

Théorème 5.5.4. *Soit X une variable aléatoire à valeurs dans $\overline{\mathbb{R}}^+$. La martingale positive généralisée $X_n = \mathbb{E}(X|\mathcal{F}_n)$ converge p.s. vers une variable aléatoire X_∞ telle que $X_\infty = \mathbb{E}(X|\mathcal{F}_\infty)$ et donc $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$. De plus, si T est un temps d'arrêt on a l'égalité $\mathbb{E}(X|\mathcal{F}_T) = X_T$.*

`Preuve:`
On prouve tout d'abord que $X_\infty = \mathbb{E}(X|\mathcal{F}_\infty)$.

• Soit $A \in \mathcal{F}_n$, d'après le théorème 5.5.2, alinéa 1 on a:

$$\int_A X_\infty \, d\mathbb{P} \leq \int_A X_n \, d\mathbb{P} = \int_A X \, d\mathbb{P}$$

et par conséquent on a $\int_A X_\infty \, d\mathbb{P} \leq \int_A X \, d\mathbb{P}$ pour $A \in \cup_n \mathcal{F}_n$ et donc pour $A \in \mathcal{F}_\infty$ ce qui prouve que $X_\infty \leq \mathbb{E}(X|\mathcal{F}_\infty)$ puisque les deux membres sont $\mathcal{F}_\infty$ mesurables.

• Pour obtenir l'inégalité dans l'autre sens, on remarque que si Z est bornée alors $\lim \mathbb{E}(Z|\mathcal{F}_n) = \mathbb{E}(Z|\mathcal{F}_\infty)$ puisque dans ce cas la martingale $\mathbb{E}(Z|\mathcal{F}_n) =$ est bornée dans L^2. On a donc pour tout $p \geq 1$:

$$X_\infty = \lim_n \mathbb{E}(X|\mathcal{F}_n) \geq \lim_n \mathbb{E}(X \wedge p|\mathcal{F}_n) = \mathbb{E}(X \wedge p|\mathcal{F}_\infty)$$

et l'on obtient la majoration souhaitée en faisant tendre p vers l'infini.

• Si T est un temps d'arrêt, il suffit de prouver que $X_n \mathbb{1}_{\{T=n\}} = \mathbb{E}(X|\mathcal{F}_T)\mathbb{1}_{\{T=n\}}$ pour $n \in \bar{\mathbb{N}}$ ce qui résulte à nouveau de la prop.5.4.8 ∎

5.6. Annexe

5.6.1. Application aux chaînes de Markov

Les fonctionnelles des chaînes de Markov fournissent une grande part des exemples concrets de martingales et surmartingales. Nous commençons par un exemple de calcul d'espérance de temps d'entrée.

Exemple

Soit $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes telles que, pour tout n, $\mathbb{P}(Y_n = 1) = p$, $\mathbb{P}(Y_n = -1) = q = 1 - p$ avec $0 < p < 1$. On pose

$$S_0 = 0, \;\; \mathcal{F}_0 = \{\Omega, \emptyset\} \text{ et pour } n \geq 1, \;\; S_n = Y_1 + \ldots + Y_n, \;\; \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$$
$$T_c = \inf(n, \; S_n = c), \quad c \in \mathbb{Z}, \quad T = T_{-a} \wedge T_b \;\; a, b \in \mathbb{N}^*$$

T est le temps de sortie de S_n de $]-a, b[$.

(i) On suppose $p > q$ (le cas $q > p$ se traite de la même façon).

On a (loi des grands nombres) $\frac{S_n}{n} \to \mathbb{E}(Y_1) = p - q > 0$ p.s. et donc $S_n \to +\infty$ p.s. Vu que S_n se déplace de ± 1, on a, pour tout $c > 0$, $\mathbb{P}(T_c < +\infty) = 1$ et a fortiori $\mathbb{P}(T < +\infty) = 1$. Cependant T n'est pas borné (sauf si $a = b = 1$). On pose

$$u(-a) = \mathbb{P}(S_T = -a) = \mathbb{P}(T_{-a} < T_b), \quad u(b) = \mathbb{P}(S_T = b) = \mathbb{P}(T_b < T_{-a}).$$

On veut calculer $u(-a)$ et $u(b)$. On a évidemment $u(-a) + u(b) = 1$.
Soit $Z_n = (\frac{q}{p})^{S_n}$, Z_n est une martingale car p.s.

$$\mathbb{E}^{\mathcal{F}_n}((\frac{q}{p})^{S_{n+1}}) = (\frac{q}{p})^{S_n}\mathbb{E}^{\mathcal{F}_n}((\frac{q}{p})^{Y_{n+1}}) = Z_n\mathbb{E}((\frac{q}{p})^{Y_{n+1}}) = Z_n(p\frac{q}{p} + q\frac{p}{q}) = Z_n.$$

On a donc $\mathbb{E}((\frac{q}{p})^{S_{T\wedge n}}) = \mathbb{E}((\frac{q}{p})^{S_0}) = 1$. Mais $Z_{T\wedge n} \to Z_T$ et $0 \leq Z_T \leq 1$ donc $\mathbb{E}(Z_{T\wedge n}) \to \mathbb{E}(Z_T)$ et $\mathbb{E}(Z_T) = 1$. (On peut aussi appliquer le cor.5.2.3). On a alors $1 = \mathbb{E}(Z_T) = (\frac{q}{p})^{-a}u(-a) + (\frac{q}{p})^b u(b)$ d'où

$$u(-a) = \frac{1 - (\frac{q}{p})^b}{(\frac{p}{q})^a - (\frac{q}{p})^b}, \quad u(b) = \frac{(\frac{p}{q})^a - 1}{(\frac{p}{q})^a - (\frac{q}{p})^b}.$$

Remarquons maintenant que $S_n - n(p-q) = \sum_{k=1}^n (Y_k - \mathbb{E}(Y_k))$ est une martingale. On a donc $\mathbb{E}(S_{T\wedge n}) = (p-q)\mathbb{E}(T \wedge n)$. Puisque $|S_{T\wedge n}| \leq \max(a,b)$, $\mathbb{E}(S_{T\wedge n}) \to \mathbb{E}(S_T)$ et puisque $T \wedge n \uparrow T$, $\mathbb{E}(T\wedge n) \uparrow \mathbb{E}(T)$, on obtient donc $(p-q)\mathbb{E}(T) = \mathbb{E}(S_T) = -au(-a) + bu(b)$ d'où

$$\mathbb{E}(T) = \frac{b((\frac{p}{q})^a - 1) - a(1 - (\frac{q}{p})^b)}{(\frac{p}{q})^a - (\frac{q}{p})^b}.$$

(ii) On suppose $p = q = \frac{1}{2}$. On sait que la marche aléatoire S_n est irréductible récurrente. Ceci implique que, pour tout $c \in \mathbb{Z}$, $\mathbb{P}(T_c < +\infty) = 1$ et donc $\mathbb{P}(T < +\infty) = 1$ mais évidemment T n'est pas borné.
Maintenant S_n est une martingale d'où $\mathbb{E}(S_{T\wedge n}) = 0$. Mais $|S_{T\wedge n}| \leq \max(a,b)$ et donc $\mathbb{E}(S_{T\wedge n}) \to \mathbb{E}(S_T)$ ce qui donne $0 = \mathbb{E}(S_T) = -au(-a) + bu(+b)$ d'où l'on tire

$$u(-a) = \frac{b}{a+b}, \quad u(b) = \frac{a}{a+b}.$$

Pour terminer calculons $\mathbb{E}(T)$. On a

$$\mathbb{E}^{\mathcal{F}_n}(S_{n+1}^2 - S_n^2) = \mathbb{E}^{\mathcal{F}_n}(Y_{n+1}(2S_n + Y_{n+1})) = S_n\mathbb{E}(Y_{n+1}) + \mathbb{E}(Y_{n+1}^2) = 1 \text{ p.s.}$$

ce qui montre que la décomposition de Doob de la sous-martingale S_n^2 est $S_n^2 = M_n + n$ ou, plus simplement, que $M_n = S_n^2 - n$ est une martingale. On a donc $\mathbb{E}(S_{T\wedge n}^2) = \mathbb{E}(T\wedge n)$. Vu que $S_{T\wedge n}^2 \leq \max(a^2, b^2)$, $\mathbb{E}(S_{T\wedge n}^2) \to \mathbb{E}(S_T^2)$ et, par convergence monotone, $\mathbb{E}(T\wedge n) \uparrow \mathbb{E}(T)$ d'où $\mathbb{E}(T) = \mathbb{E}(S_T^2) = a^2u(-a) + b^2u(b)$ et $\mathbb{E}(T) = ab$.

Compléments de théorie du potentiel

Le résultat suivant montre les liens entre martingales et chaînes de Markov.

Proposition 5.6.1. *On rappelle qu'une fonction f définie sur E et à valeurs dans $[0, +\infty]$ est dite surharmonique (resp. harmonique) si $Qf \leq f$ (resp $=$). Soit $f \in \mathcal{E}^+$ surharmonique (resp. harmonique). Alors $Y_n = f(X_n)$ est, pour toute loi $\mathbb{P}_\mu$, une surmartingale (resp. une martingale) généralisée positive.*

`Preuve:`
Soit par exemple $f \in \mathcal{E}^+$ harmonique. On a alors $\mathbb{E}_\mu(Y_{n+1}|\mathcal{F}_n) = Qf(X_n) = f(X_n) = Y_n$, $\mathbb{P}_\mu$ p.s. ∎

Proposition 5.6.2. *Soit X une chaîne de Markov irréductible. Si X est récurrente alors toute fonction excessive est constante.*

`Preuve:`
Soit f une fonction excessive. Alors $f(X_n)$ est, pour toute loi $\mathbb{P}_x$, une surmartingale positive généralisée et pour tous $x, y \in E$, $\mathbb{P}_x(T_y < \infty) = 1$. D'après le cor.5.5.3, pour tous $x, y \in E$, $f(x) = \mathbb{E}_x(f(X_0)) \geq \mathbb{E}_x(f(X_{T_y})) = f(y)$ d'où, x, y étant arbitraires, $f(x) = f(y)$. (Ce résultat avait déjà été obtenu en 4.9.2) ∎

Définition 5.6.3. *Soit $B \subset E$. On pose:*

$$\pi_B(x) = \mathbb{P}_x(T_B < \infty), \quad h_B(x) = \mathbb{P}_x(N_B = \infty)$$

Proposition 5.6.4. *La fonction π_B est sur-harmonique, la suite (décroissante) $Q^n \pi_B(x)$ converge vers la fonction harmonique $h_B(x)$. De plus pour tout $x \in E$, la $\mathbb{P}_x$ surmartingale $\pi_B(X_n)$ et la $\mathbb{P}_x$ martingale $h_B(X_n)$ convergent $\mathbb{P}_x$ p.s. vers $\mathbb{1}_{\{N_B = \infty\}}$*

`Preuve:`
On pose $T_B^k = \inf\{n \geq k\ ;\ X_n \in B\}$, alors $\mathbb{1}_{\{T_B^k < \infty\}} = \mathbb{1}_{\{T_B < \infty\}} \circ \theta_k$ et:

$$\begin{aligned} Q^k \pi_B(x) &= \mathbb{E}_x(\pi_B \circ X_k) = \mathbb{E}_x\Big(\mathbb{P}_{X_k}(T_B < \infty)\Big) \\ &= \mathbb{E}_x\Big(\mathbb{E}_x(\mathbb{1}_{\{T_B < \infty\}} \circ \theta_k | \mathcal{F}_k)\Big) = \mathbb{P}_x(T_B^k < \infty) \\ &\leq \pi_B(x) \end{aligned}$$

Ceci prouve que π_B est sur-harmonique et que

$$\lim_n \downarrow Q^n \pi_B(x) = \lim_n \downarrow \mathbb{P}_x(T_B^n < \infty) = \mathbb{P}_x(\cap_n (T_B^n < \infty)) = h_B(x)$$

car les T_B^n forment une suite croissante. On en déduit immédiatement que h_B est harmonique puisque $Qh_B = Q(\lim_n \downarrow Q^n \pi_B) = \lim_n \downarrow Q^{n+1} \pi_B = h_B$. La suite $\pi_B(X_n)$ est une surmartingale positive et $h_B(X_n)$ est une martingale positive, elles convergent donc $\mathbb{P}_x$ p.s. et ces suites étant bornées, elles convergent dans L^1. Or on a: $\mathbb{1}_{\{N_B = \infty\}} \circ \theta_n = \mathbb{1}_{\{N_B = \infty\}}$ et donc

$$h_B(X_n) = \mathbb{E}_{X_n}(\mathbb{1}_{\{N_B = \infty\}}) = \mathbb{E}_x(\mathbb{1}_{\{N_B = \infty\}} \circ \theta_n | \mathcal{F}_n) = \mathbb{E}_x(\mathbb{1}_{\{N_B = \infty\}} | \mathcal{F}_n)$$

Le théorème (5.5.4) implique que $h_B(X_n)$ converge $\mathbb{P}_x$ p.s. vers $Y = \mathbb{1}_{\{N_B=\infty\}}$. Si Z est la limite de $\pi_B(X_n)$ on a bien sûr $Z \geq Y$ et le théorème de Lebesgue implique que

$$\mathbb{E}_x(Z) = \lim \mathbb{E}_x(\pi_B(X_n)) = \lim \mathbf{Q}^n \pi_B(x) = h_B(x) = \mathbb{E}_x(Y)$$

et donc $Z = Y$, $\mathbb{P}_x$ p.s. ∎

Cette proposition a d'importantes conséquences et permet de retrouver une bonne partie des résultats obtenus dans la classification des états d'une chaîne de Markov.

Proposition 5.6.5. *Soit $x \in E$. Alors $\mathbb{P}_x(N_x = \infty)$ ne prend que les valeurs 0 ou 1.*

`Preuve:`
La proposition 5.6.4 implique que la suite $h_x(X_n)\mathbb{1}_{\{N_x=\infty\}}$ converge $\mathbb{P}_x$ p.s. vers $\mathbb{1}_{\{N_x=\infty\}}$. Mais cette suite est égale à $h_x(x)$ pour une infinité de valeurs de n et donc $h_x(x) = \mathbb{1}_{\{N_x=\infty\}}$, $\mathbb{P}_x$ p.s. On en déduit que si $h_x(x) > 0$ alors $h_x(x) = 1$.Ce résultat avait déjà été obtenu dans le théorème 4.5.4. ∎

Proposition 5.6.6. *Soit x et $y \in E$ avec $\mathbb{P}_x(N_y = \infty) = 1$. Alors pour tout point $z \in E$ il n'y a que deux cas possibles:*

- $\mathbb{P}_y(T_z < \infty) = \mathbb{P}_y(N_z = \infty) = \mathbb{P}_x(N_z = \infty) = 1.$
- $\mathbb{P}_y(T_z < \infty) = \mathbb{P}_y(N_z = \infty) = \mathbb{P}_x(N_z = \infty) = 0.$

`Preuve:`
La proposition 5.6.4 implique que les suites $h_z(X_n)$ et $\pi_z(X_n)$ convergent $\mathbb{P}_x$ p.s. vers $\mathbb{1}_{\{N_z=\infty\}}$. Mais, $\mathbb{P}_x$ p.s., pour une infinité de valeurs de n, la première suite est égale à $h_z(y)$ et la seconde à $\pi_z(y)$ et donc $\mathbb{P}_y(T_z < \infty) = \mathbb{P}_y(N_z = \infty) = \mathbb{1}_{\{N_z=\infty\}}$ $\mathbb{P}_x$ p.s. ∎

Corollaire 5.6.7. *Si $\mathbb{P}_x(N_x = \infty) = 1$, alors pour tout z tel que $\mathbb{P}_x(T_z < \infty) > 0$ on a $\mathbb{P}_x(T_z < \infty) = \mathbb{P}_x(N_z = \infty) = 1$. (Autrement dit, si x est récurrent et $x \to z$ alors $\mathbb{P}_x(N_z = \infty) = 1$ et z est récurrent d'après le principe du maximum.)*

`Preuve:`
Il suffit de prendre $x = y$ dans la proposition précédente. Ce résultat avait déjà été obtenu en 4.5.12. ∎

Proposition 5.6.8. *Soient A et B contenus dans E et $\delta > 0$ tels que pour tout $x \in A$, $\mathbb{P}_x(T_B < +\infty) \geq \delta$. Alors, pour tout $x \in E$, $(N_A = \infty) \subset (N_B = \infty)$, P_x p.s.*

`Preuve:`
Soit $R = \inf\{n \geq T_A \; ; \; X_n \in B\}$, alors $R \geq T_B$ et sur $(T_A < \infty)$ on a $R = T_A + T_B \circ \theta_{T_A}$. Par application de Markov fort on obtient:

$$\mathbb{P}_x(R < \infty) = \mathbb{P}_x((T_A < \infty) \cap \theta_{T_A}^{-1}(T_B < \infty)) = \mathbb{E}_x(\mathbb{1}_{\{T_A<\infty\}} \mathbb{E}_{X_{T_A}}(T_B < \infty))) \geq \delta \mathbb{P}_x(T_A < \infty)$$

En appliquant la proposition 5.6.4 et le fait que $\mathbb{P}_x(T_B < \infty) \geq \mathbb{P}_x(R < \infty)$ on obtient que $\mathbb{1}_{\{N_B=\infty\}} \geq \delta \mathbb{1}_{\{N_A=\infty\}}$ $\mathbb{P}_x$ p.s. d'où le résultat. ∎

5.6.2. Etude des sous-martingales

On commence par une généralisation de l'étude des sous martingales entreprise dans 5.4.2.

Définition 5.6.9. *On dit qu'un processus X_n est à accroissements bornés, s'il existe un réel $K < \infty$ tel que $|X_n - X_{n-1}| \leq K$, p.s. pour tout $n \geq 1$.*

Théorème 5.6.10. *Soit X_n une sous-martingale, et $X_n = M_n + A_n$ sa décomposition de Doob.*

1. *Si X_n est positive alors X_n converge p.s. sur l'ensemble $\{A_\infty < \infty\}$ vers une limite p.s finie.*

2. *Si X_n est à accroissements bornés, alors on a l'inclusion:*

$$\{\sup X_n < +\infty\} \subset (\{X_n \text{ converge dans } \mathbb{R}\} \cap \{A_\infty < +\infty\}) \text{ p.s.}$$

   (ce qui implique en fait l'égalité p.s. de ces deux ensembles.)

En particulier, pour une sous martingale positive à accroissements bornés on a

$$\{X_n \text{ converge dans } \mathbb{R}\} = \{A_\infty < +\infty\} \text{ p.s.}$$

`Preuve:`
• Pour un entier $p > 0$ on pose $T_p = \inf(n\,,\, A_{n+1} > p)$. T_p est un temps d'arrêt, $A_{T_p \wedge n} \leq p$ et $T_p = +\infty$ sur $\{A_\infty \leq p\}$. Alors $Z_n = M_{T_p \wedge n}$ est une martingale telle que $Z_n^- \leq p$, elle est par conséquent bornée dans L^1 et converge donc p.s. On en déduit la convergence p.s. de Z_n et donc celle de X_n sur $\{T_p = +\infty\}$ et enfin celle de X_n sur $\cup_p \{T_p = +\infty\} = \{A_\infty < +\infty\}$.

• On peut supposer $X_0 = 0$ en considérant la sous martingale $X_n - X_0$ qui a le même processus croissant. On pose, pour un entier $p > 0$, $T_p = \inf(n \geq 1,\ X_n > p)$. On a $X_{T_p \wedge n} = X_n \mathbb{1}_{\{T_p > n\}} + X_{T_p} \mathbb{1}_{\{T_p \leq n\}} \leq 2p + K$ donc, $\mathbb{E}(X^+_{T_p \wedge n}) \leq 2p + K$. La sous martingale $X_{T_p \wedge n}$ est donc bornée dans L^1 et par conséquent elle converge p.s.. On en déduit que X_n converge p.s sur $\{T_p = +\infty\}$. Par ailleurs $\mathbb{E}(A_{T_p \wedge n}) = \mathbb{E}(X_{T_p \wedge n}) \leq \mathbb{E}|X_{T_p \wedge n}|$ donc $C = \sup_n \mathbb{E}(A_{T_p \wedge n}) < +\infty$ et, par limite croissante, $\mathbb{E}(A_{T_p}) \leq C$ d'où $A_{T_p} < +\infty$ p.s. et $A_\infty < +\infty$ p.s. sur $\{T_p = +\infty\}$. On conclut en remarquant que $\{\sup X_n < +\infty\} = \cup_p \{T_p = +\infty\}$ ∎

Corollaire 5.6.11. *Soit Y_n, $n \geq 0$ un processus positif intégrable. On pose pour $n \geq 1$:*

$$X_n = Y_1 + \ldots Y_n, \qquad \widetilde{X_n} = \mathbb{E}(Y_1|\mathcal{F}_0) + \ldots + \mathbb{E}(Y_n|\mathcal{F}_{n-1})$$

et $X_0 = Y_0$, $\widetilde{X_0} = 0$. Alors on a $\{\widetilde{X_\infty} < \infty\} \subset \{X_\infty < \infty\}$ avec égalité si le processus Y_n est borné par une constante $K < \infty$.

`Preuve:`
Il suffit de remarquer que X_n est une sous martingale positive, dont le processus croissant est $\widetilde{X_n}$ et d'appliquer le théorème 5.6.10 ∎

On en déduit facilement le lemme de Borel-Cantelli conditionnel

Proposition 5.6.12. *Soit $(B_n)_{n\geq 1}$ une suite d'événements avec $B_n \in \mathcal{F}_n$. On a p.s.*

$$\{\sum_{n\geq 1} \mathbb{1}_{\{B_n\}} < +\infty\} = \{\sum_{n\geq 1} \mathbb{P}(B_n|\mathcal{F}_{n-1}) < +\infty\}.$$

On peut aussi préciser le théorème 5.3.8 lorsque les sauts de X_n sont bornés.

Corollaire 5.6.13. *Soit X_n une martingale de carré intégrable à accroissements bornés, et soit A_n le processus croissant dans la décomposition de Doob de X_n^2. Alors p.s.*

$$\{A_\infty < +\infty\} = \{X_n \text{ converge dans } \mathbb{R}\}, \qquad \{A_\infty = \infty\} = \{\overline{\lim}\, X_n = +\infty \ ,\ \underline{\lim}\, X_n = -\infty\}$$

`Preuve:`
En appliquant le théorème 5.6.10 aux sous-martingales X_n et $-X_n$, on obtient l'inclusion:

$$\{X_n \text{ ne converge pas}\} \subset \{\sup X_n = +\infty \ ,\ \inf X_n = -\infty\}$$

On sait déjà que $\{A_\infty < +\infty\} \subset \{X_n \text{ converge}\}$. Pour obtenir l'inclusion inverse, on ne peut appliquer directement le second alinéa du théorème 5.6.10 car X_n^2 n'est peut-être pas à accroissements bornés. Par contre on peut reprendre la preuve de ce second alinéa en considérant la sous martingale X_n^2. En effet, avec les mêmes notations (en posant $T_p = \inf(n \geq 0,\ X_n^2 > a))$ on aura :

$$X_{T_p\wedge n}^2 = X_n^2 \mathbb{1}_{\{T_p>n\}} + X_{T_p}^2 \mathbb{1}_{\{T_p\leq n\}} \leq 2p + |X_{T_p}^2 - X_{T_p-1}^2| \mathbb{1}_{\{T_p<+\infty\}}$$

En utilisant la majoration $|y^2-x^2| \leq |y-x|^2+2|x||y-x|$ on obtient que $|X_{T_p}^2 - X_{T_p-1}^2| \leq K^2 + 2K\sqrt{p}$ et on peut alors recopier la fin de la preuve ∎

Remarque. On ne peut pas espérer un résultat aussi fort sans hypothèse sur les sauts de X_n. Soit $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes telles que $\mathbb{P}(Y_n = a_n) = p_n$, $\mathbb{P}(Y_n = -a_n) = p_n$ et $\mathbb{P}(Y_n = 0) = 1 - 2p_n$ avec $a_n > 0$, $0 < p_n < \frac{1}{2}$. $X_n = \sum_{k=1}^n Y_k$ est une martingale de carré intégrable et $A_n = \sum_{k=1}^n \mathbb{E}(Y_k^2) = 2\sum_{k=1}^n p_n a_n^2$. On a $\sum_n \mathbb{P}(Y_n \neq 0) = 2\sum_n p_n$. Donc si $\sum_n p_n < +\infty$, d'après Borel-Cantelli, $Y_n = 0$ à partir d'un certain rang p.s. et $X_n = \sum_{k=1}^n Y_k$ converge p.s. alors qu'on peut avoir $A_\infty = 2\sum p_n a_n^2 = +\infty$ (prendre par exemple $p_n = n^{-2}$ et $a_n = n$).

5.6.3. Suites de v.a.r. indépendantes

Soit $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes, intégrables et centrées. On pose

$$S_0 = 0,\ \ \mathcal{F}_0 = \{\Omega, \emptyset\} \text{ et pour } n \geq 1,\ S_n = Y_1 + \ldots + Y_n,\ \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n).$$

Alors S_n est une martingale. Si, pour tout n, $\mathbb{E}(Y_n^2) < +\infty$, S_n est une martingale de carré intégrable. Dans ce cas,

$$\mathbb{E}((S_{n+1}^2 - S_n^2 | \mathcal{F}_n)) = \mathbb{E}((Y_{n+1}(2S_n + Y_{n+1}))|\mathcal{F}_n) = 2S_n\mathbb{E}(Y_{n+1}) + \mathbb{E}(Y_{n+1}^2) = \mathbb{E}(Y_{n+1}^2) \text{ p.s.}$$

et donc, pour $n \geq 1$, le processus croissant associé à la sous martingale S_n^2 est égal au processus déterministe $\sum_{k=1}^n \mathbb{E}(Y_k^2)$.

Lois des grands nombres.

Proposition 5.6.14. *Soient $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes de carré intégrable et $b_n > 0$, $b_n \to +\infty$. On suppose que $\frac{1}{b_n} \sum_{k=1}^n \mathbb{E}(Y_k) \to_n m$. Alors,*

$$\text{si } \sum \frac{Var(Y_n)}{b_n^2} < +\infty, \quad \frac{S_n}{b_n} \to m \text{ p.s.}$$

`Preuve:`
On peut supposer $\mathbb{E}(Y_k) \equiv 0$ et alors $m = 0$. La martingale $Z_n = \sum_{k=1}^n \frac{Y_k}{b_k}$ est bornée dans L^2 et donc converge p.s. D'après le lemme de Kronecker (lem.5.3.7), $\frac{S_n}{b_n} \to 0$ p.s.
∎

Corollaire 5.6.15. *Si $Y_1, \ldots, Y_n, \ldots$ est une suite de v.a.r. indépendantes avec, pour tout k, $\mathbb{E}(Y_k) = m$ et $\sup_k \mathbb{E}(Y_k^2) = M < +\infty$, alors on a $\frac{S_n}{n} \to m$ p.s. Si de plus les Y_n sont centrées alors pour tout $\delta > 1/2$, on a $\frac{S_n}{n^\delta} \to_n 0$ p.s.*

On va maintenant établir la loi forte des grands nombres, sans supposer l'existence d'un moment d'ordre 2.

Théorème 5.6.16. *Soit $(Y_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi avec $E|Y_1| < +\infty$. Alors $\frac{S_n}{n} \to_n \mathbb{E}(Y_1)$ p.s. et dans L^1.*

`Preuve:`
On peut supposer $\mathbb{E}(Y_1) = 0$. On remarque que la suite $M_n = S_n/n$ est uniformément intégrable puisque:

- M_n est bornée dans L^1 car $\mathbb{E}(|M_n|) \leq (\sum_{k=1}^n \mathbb{E}(|Y_k|))/n = \mathbb{E}(|Y_1|)$
- M_n est équi-continue car $\int_A |M_n| \, d\mathbb{P} \leq (\sum_{k=1}^n \int_A |Y_k| \, d\mathbb{P})/n \leq \sup_k \int_A |Y_k| \, d\mathbb{P}$ et il est évident que la suite Y_n est uniformément intégrable donc équicontinue puisque $\int_{|Y_k|>a} |Y_k| \, d\mathbb{P}$ ne dépend pas de k.

Il suffit donc de prouver la convergence p.s. pour avoir aussi la convergence L^1. On pose $Z_n = Y_n \mathbb{1}_{\{|Y_n| \leq n\}}$. Nous allons tout d'abord montrer que les trois propriétés suivantes permettent d'obtenir le résultat cherché.

$$(1) \sum_n \mathbb{P}(Y_n \neq Z_n) < \infty, \qquad (2) \lim_n \mathbb{E}(Z_n) = 0, \qquad (3) \sum_n \mathbb{E}(Z_n^2)/n^2 < \infty$$

En effet (2) et (3) et la proposition 5.6.14 permettent d'affirmer que $\frac{1}{n}\sum_{k=1}^{n} Z_k \to_n 0$ p.s. De plus d'après le lemme de Borel Cantelli (1) implique que p.s. on a $Y_n = Z_n$ à partir d'un certain rang (aléatoire) et l'on obtient donc le résultat voulu. Il reste à prouver ces trois propriétés. Soit Y une variable aléatoire ayant la loi commune des Y_n.

(1) On a $\sum_n \mathbb{P}(Y_n \neq Z_n) = \sum_n \mathbb{P}(|Y_n| > n) = \sum_n \mathbb{P}(|Y| > n) \leq \mathbb{E}(|Y|)$

(2) On a $\mathbb{E}(Z_n) = \mathbb{E}(Y_n \mathbb{1}_{\{|Y_n| \leq n\}}) = \mathbb{E}(Y \mathbb{1}_{\{|Y| \leq n\}})$ et cette dernière quantité tend vers $\mathbb{E}(Y) = 0$ par le théorème de Lebesgue.

(3) On a $\mathbb{E}(Z_n^2) = \mathbb{E}(Y_n^2 \mathbb{1}_{\{|Y_n| \leq n\}}) = \mathbb{E}(Y^2 \mathbb{1}_{\{|Y| \leq n\}})$ et d'autre part on décompose l'ensemble $\{|Y| \leq n\}$ pour obtenir $|Y| \mathbb{1}_{\{|Y| \leq n\}} = |Y| \sum_{k=1}^{n} \mathbb{1}_{\{k-1 < |Y| \leq k\}}$ et l'on écrit:

$$\begin{aligned}
\sum_{n \geq 1} \mathbb{E}(Z_n^2)/n^2 &= \mathbb{E}(|Y| \sum_{n \geq 1} \frac{1}{n^2} (\sum_{k=1}^{n} |Y| \mathbb{1}_{\{k-1 < |Y| \leq k\}})) \\
&= \mathbb{E}(|Y| \sum_{k \geq 1} (\sum_{n \geq k} \frac{1}{n^2}) |Y| \mathbb{1}_{\{k-1 < |Y| \leq k\}}) \\
&\leq \mathbb{E}(|Y| \sum_{k \geq 1} \frac{2}{k} |Y| \mathbb{1}_{\{k-1 < |Y| \leq k\}}) \\
&\leq 2\mathbb{E}(|Y| \sum_{k \geq 1} \mathbb{1}_{\{k-1 < |Y| \leq k\}}) = 2\mathbb{E}(|Y|) < \infty \quad \blacksquare
\end{aligned}$$

Séries de variables aléatoires indépendantes.

Théorème 5.6.17. *Soit $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes de carré intégrable.*

1. *Si les séries $\sum \mathbb{E}(Y_n)$ et $\sum$ Var(Y_n) convergent, S_n converge p.s. et dans L^2.*
2. *S'il existe $M \in \mathbb{R}$ tel que, pour tout n, $|Y_n| \leq M$ p.s. et si S_n converge p.s., les séries $\sum \mathbb{E}(Y_n)$ et $\sum$ Var(Y_n) convergent.*

`Preuve:`
A. On suppose $\mathbb{E}(Y_k) \equiv 0$. Alors S_n est une martingale de carré intégrable. Donc si $\sum_n \text{Var}(Y_n) < +\infty$, S_n est bornée dans L^2 et donc (th.5.3.4) S_n converge p.s. et dans L^2. Réciproquement si $|Y_n| \leq M$ p.s., d'après le th.5.6.13, S_n converge p.s. implique que le processus croissant dans la décomposition de Doob de X_n^2 soit tel que $A_\infty < \infty$ p.s. Or $A_\infty = \sum_n \text{Var}(Y_n)$.

B. Cas général. Ecrivant $\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n}(Y_k - \mathbb{E}(Y_k)) + \sum_{k=1}^{n} \mathbb{E}(Y_k)$, (1) est une conséquence immédiate de A. Le point (ii) est un peu plus délicat. Soient $Y_1', Y_1'', \ldots, Y_n', Y_n'', \ldots$ une suite de v.a.r. toutes indépendantes définies sur $(\Omega, \mathcal{G}, \mathbb{Q})$ telles que, pour tout n, loi de Y_n'=loi de Y_n''=loi de Y_n. On pose

$$S_n' = \sum_{k=1}^{n} Y_k', \quad S_n'' = \sum_{k=1}^{n} Y_k'', \quad Z_n = Y_n' - Y_n''.$$

On a $\mathbb{E}_{\mathbb{Q}}(Z_n) = 0$, $\text{Var}_{\mathbb{Q}}(Z_n) = \text{Var}_{\mathbb{Q}}(Y'_n) + \text{Var}_{\mathbb{Q}}(Y''_n) = 2\text{Var}(Y_n)$. Vu que les processus $((S'_n)_{n\geq 1}, \mathbb{Q})$ et $((S''_n)_{n\geq 1}, \mathbb{Q})$ ont même loi que $((S_n)_{n\geq 1}, \mathbb{P})$, le lem.5.3.3 implique que S'_n et S''_n convergent $\mathbb{Q}$ p.s. et donc aussi $\sum_{k=1}^{n} Z_k = S'_n - S''_n$. On a donc, vu A puisque les Z_n sont centrées et p.s. uniformément bornées, $\sum_n \text{Var}_{\mathbb{Q}}(Z_n) < +\infty$ d'où $\sum_n \text{Var}(Y_n) < +\infty$. D'après A, $\sum_{k=1}^{n}(Y_k - \mathbb{E}(Y_k))$ converge p.s. et, puisque $\sum_{k=1}^{n} Y_k$ converge, $\sum_{k=1}^{n} \mathbb{E}(Y_k)$ converge ∎

Remarque. On ne peut pas espérer (2) sans hypothèses supplémentaires. Voir l'exemple de la remarque suivant 5.6.13.

Critère des trois séries.

Théorème 5.6.18. *Soient $Y_1, \ldots, Y_n, \ldots$ une suite de v.a.r. indépendantes et $K > 0$. On pose $Y_n^K = Y_n \mathbb{1}_{\{|Y_n| \leq K\}}$. Il y a équivalence entre*

1. *$\sum_{k=1}^{n} Y_k$ converge p.s.*
2. *Les séries $\sum \mathbb{P}(|Y_n| > K)$, $\sum \mathbb{E}(Y_n^K)$, $\sum$ Var(Y_n^K) convergent.*

`Preuve:`
(1) Supposons que $\sum_n Y_n$ converge p.s. Alors $\sum \mathbb{P}(|Y_n| > K) < +\infty$ car, si $\sum \mathbb{P}(|Y_n| > K) = +\infty$, on a p.s. (prop.2.2.11) $|Y_n| > K$ infiniment souvent et S_n diverge p.s. Alors la convergence de $\sum \mathbb{P}(|Y_n| > K)$ implique (prop.2.2.11) que p.s. $|Y_n| > K$ n'a lieu qu'un nombre fini de fois. Les séries $\sum Y_n$ et $\sum Y_n^K$ sont donc p.s. de même nature et $\sum Y_n^K$ converge p.s. Puisque $|Y_n^K| \leq K$, on peut appliquer le th.5.6.17 et $\sum \mathbb{E}(Y_n^K)$ et $\sum \text{Var}(Y_n^K)$ convergent.

(2) Supposons que les trois séries convergent. Vu le th.5.6.17, $\sum Y_n^K$ converge p.s. et, comme ci-dessus, la convergence de $\sum \mathbb{P}(|Y_n| > K)$ implique que les séries $\sum Y_n$ et $\sum Y_n^K$ sont p.s. de même nature. Donc $\sum_n Y_n$ converge p.s. ∎

Bibliographie

[1] P.Barbe, M.Ledoux: *Probabilité*, Berlin, Paris, 1999

[2] P.Billingsley: *Probability and Measure*, Wiley and sons, New-York, 1979

[3] N.Bouleau: *Processus stochastiques et applications*, Hermann, Paris, 1988

[4] P.Bremaud: *Introduction aux probabilités. Modélisation des phénomènes aléatoires*, Springer-verlag, New-York, 1984

[5] L.Breiman: *Probability*, Addison-Wesley, Reading, 1968

[6] M.Cottrel, Ch.Duhamel, V.Genon-Catalot: *Exercices de Probabilités*, Berlin, Paris, 1980

[7] D.Dacunha-Castelle, M.Duflo: *Probabilités et Statistiques*, Tome 1: Problèmes à temps fixe, Masson, Paris, 1982

[8] D.Dacunha-Castelle, M.Duflo: *Probabilités et Statistiques*, Tome 2: Problèmes à temps mobile, Masson, Paris, 1983

[9] D.Dacunha-Castelle, M.Duflo: *Exercices de Probabilités et Statistiques*, Tome 1: Problèmes à temps fixe, Masson, Paris, 1982

[10] D.Dacunha-Castelle, M.Duflo, V.Genon-Catalot: *Exercices de Probabilités et Statistiques*, Tome 2: Problèmes à temps mobile, Masson, Paris, 1983

[11] D.Dacunha-Castelle, D.Revuz, M.Schreiber: *Recueil de Problèmes de Calcul des Probabilités*, Masson, Paris, 1970

[12] C.Dellacherie, P-A.Meyer: *Probabilités et Potentiel*, Vol.2, Hermann, Paris, 1980

[13] R.Durrett: *Probability: Theory and Examples*, Wadsworth and Brooks, Pacific Grove, 1991

[14] P.Hoel, S.Port, C.Stone: *Introduction to Stochastic Processes*, Houghton Mifflin, Boston, 1972

[15] G.Letac: *Problèmes de probabilité*, Collection SUP, PUF, Paris, 1970

[16] G. Letac: *Intégration et probabilités, analyse de Fourier et analyse spectrale. Exercices*, Masson, 1997

[17] L.Mazliak, P.Priouret, P.Baldi: *Martingales et chaînes de Markov*, Hermann, Paris, 1998

[18] J.Neveu: *Bases mathématiques des probabilités*, Masson, Paris, 1964

[19] J.Neveu: *Martingales à temps discret*, Masson, Paris, 1972

[20] J.R.Norris: *Markov chains*, Cambridge University Press, 1997

[21] D.Revuz: *Mesures et intégration*, Hermann, Paris, 1994

[22] D.Revuz: *Probabilités*, Hermann, Paris, 1997

[23] W.Rudin: *Analyse réel et complexe*, Masson, Paris, 1995

[24] A.N.Shiriyaev: *Probability*, Springer-Verlag, New-York, 1984

[25] D.Williams: *Probability with Martingales*, Cambridge Math.Textbooks, Cambridge, 1991